

KÖZGAZDASÁGI GONDOLATOK A KÖZSZFÉRA ADATAIHOZ VALÓ HOZZÁFÉRÉSRŐL

Cseres-Gergely Zsombor és Csorba Gergely

A tanulmány az Eötvös Károly Intézet „Hozzáférés a nemzeti adatvagyonhoz” c. kutatásának részeként készült. A kutatás különböző pontjain nyújtott segítő megjegyzésekért, tanácsokért és vitákért köszönetünket szeretnénk kifejezni az MTA Közgazdaságtudományi Intézete munkatársainak, különösen Kézdi Gábornak és Köllő Jánosnak. Igen hasznos kutató-asszisztensi segítséget kaptunk Bálint Mónikától és Nárai Csillától, akiket külön köszönet illet. Végül szeretnénk megköszönni az OMSZ, MGSZ, KSH és APEH munkatársainak az együttműködést. A tanulmányban fennmaradó esetleges hibákért a felelősség természetesen csak minket terhel.

Bevezetés

Jelen kutatás célja közgazdasági eszközökkel vizsgálni közérdekű adatok szabad hozzáférésűvé tételének hatását, és ajánlásokat tenni annak módjára, szerkezetére, illetve mértékére. Ennek érdekében segítségül hívunk közgazdasági elméleti eredményeket, és áttekintjük a hazai adathozzáférési gyakorlatot, valamint nemzetközi példákat.

Adatnak a köznyelvben rendszerint száraz, fragmentált tudáselemet nevezünk – a számítógépek korában talán az „adatrekord” szó jeleníti meg legjobban ennek a fogalomnak a jelentését. Ezzel szemben a „tudás”-hoz közelebb áll az *ismeret* fogalma, amely a legtöbbször az adaton túlmutató, az adatból, az adatok halmazából levont összefüggéseket jelenít meg, és egyértelműen összetettebb konstrukcióként jelenik meg. A két értelmezés – amelyet a későbbiekben az Avtv. szellemének megfelelően felváltva használunk – természetesen nem zárja ki egymást: az elemi adat nemcsak lehet az összetettebb ismeret alapja, de az ismeret a legtöbb fontos esetben csak elemi adatok értelmezésével szerezhető meg.

Az Avtv. azért támogatja az állami intézményekre vonatkozó, vagy általuk kezelt adatok szabad hozzáférhetőségét, hogy az azokban megtalálható ismeretek hasznosuljanak. A törvény értelmében az adat és ismeret szinonima, mintegy sugallva, hogy az adat egyben közvetlen ismeret is. Számos olyan ismeret van azonban, amely közvetett módon nem szerezhető meg, csak olyan adatok speciális feldolgozásával, amelyek önmagukban a vizsgált kérdésben nem informatívak. A jogi szabályozás sehol nem teszi nyilvánvalóvá azt a különbséget, ami a két adattípus között van, és amihez az adathozzáférés gyakorlata is idomult: sok esetben csak azokat az adatokat minősítik hasznosnak, amiknek látszanak a közvetlen felhasználási lehetőségei.

4.5. Az állami adatkezelés jelenlegi formájához vezető út

Az állam és intézményei kezelésében levő adatok keletkezésének oka eredendően nem az állampolgárok tájékoztatása, azaz a konkrét ügyön túlmutató ismeretek létrehozása, hanem az, hogy nyilvántartási adatként szolgálják az adminisztrációt, lehetővé tegyék annak zavartalan működését. Az ilyen adatok feldolgozásának egyetlen módja hosszú időn keresztül a felsőbb szintek felé tett jelentésekben megjelenő összesítés volt, a felhasználással nyerhető általános következtetéseket pedig főként történészek vonták le a keletkezés után jóval később. Az állam által felhalmozott ismeretek tehát elsősorban egyedi és azonnali következtetés levonását tették lehetővé a saját maga számára, például, hogy az adófizető eleget tett-e adófizetési kötelezettségének.

Az idők folyamán két alapvető változás következett be az állam által gyűjtött adatok felhasználása terén, miután világossá vált: az ügyintézés során gyűjtött adatok azon túlmutató célra is használhatók. A világ azon részén, ahol az angolszász típusú polgári fejlődés dominánssá vált, az állam egyre hangsúlyosabban szolgáltatóként jelenik meg, és nem úgy, mint aki polgárai helyett kíván ügyeket intézni.

Az első fontos változás tehát az, hogy állam információszerző és -közlő szerepe alapvetően módosult. Nem elég az eseti ügyintézés számára elengedhetetlen információkat gyűjtenie, hanem figyelembe kell vennie polgárai igényét és megkísérelnie kielégíteni azokat akár korábban nem létező intézmények létrehozásával is. Ezt a szerepet hangsúlyozza a magyar Alkotmány és az Avtv. is. Manapság számos olyan tájékoztatási rendszert ismerünk, a statisztikától a különféle nyilvántartásokig, amelyeknek fontos (ha nem is egyetlen) célja az, hogy hiteles, könnyen elérhető információforrásként az állampolgárok rendelkezésére álljanak.

Ezzel párhuzamosan fontos hangsúlyozni az állam polgáraival szembeni szolgáltató szerepét. Érdektelen tehát, hogy az adatokat közvetlenül az állam vagy polgárai

használják fel, az egyik vagy másik csatornán történő felhasználás esetében ugyanolyan megfontolásokat kell alkalmaznunk, hiszen mindkét esetben az állampolgárok jólétének növelése a végcél. Egyedi megfontolást igényel azonban annak eldöntése, hogy az egyes esetekben a költségek és hasznok miként oszlanak meg az adatgyűjtések finanszírozói és felhasználói, illetve azok egyes csoportjai között.

A közvetlen információhasználók körének megnövekedése mellett a második fő változás minőségi, aminek háttérében egyfelől a statisztika, másfelől a számítástechnika fejlődése áll. A statisztika az adatok egyedi felhasználása mellett lehetőséget teremt arra, hogy az egyes adatok partikularitásától elvonatkoztatva olyan következtetéseket vonjunk le, amely nem egyedekre, hanem összefüggésekre vonatkoznak. E következtetések hasonlítanak a történészek által levontakra, amelyek összefüggést kerestek például a Napóleon által Oroszország ellen indított háború sikertelensége és a hadi logisztika közötti összefüggésekre. Noha az egyedi információk egy-egy feladat ellátásához igen fontosak, az általános következtetések sorsdöntők lehetnek. Ha az ellátási igény soha nem jut el a hadtáp-tiszthez időben, a katonák éhezni fognak, a választott logisztikai rendszer csődöt mond – egy „kicsit” tökéletlen rendszer azonban még kielégítő lehet. Ha azonban a hadvezér nem ismeri az egyes ellátási rendszerek erőforrásigényét és hatását, így nem a megfelelő stratégiát választja, hadjárata egészében van kudarca ítéelve, a részegységek működésétől függetlenül.

A statisztikai elmélet fejlődése önmagában kevés az egyeditől elvonatkoztatott következtetések levonására, vagy ahogy ezek után nevezzük, a kutatási célú adatfelhasználás kialakulására. Tudjuk, hogy azok az országok (és vállalatok), amelyek ezt megtehették, alkalmaztak viszonylag összetett statisztikai módszereket akkor is, amikor csak emberi munkaerőt vehettek igénybe azok működtetésére. Az adatok hasznosulását legalább ennyire befolyásoló tényező az, hogy az előállított következtetések az eseményeket milyen hosszú idő elteltével követik. A számítástechnika fejlődése volt az, amely az adatok felhasználásának lehetőségeit újradefiniálta azáltal, hogy az adatok feldolgozásához szükséges időt drámai módon

lerövidítette. Sok esetben ez azt jelentette, hogy korábban reménytelen feladatok váltak megvalósíthatóvá. Ma már nagy mennyiségű adatból gyakorlatilag azonnal, vagy igen kis késéssel vonjunk le összetett, a közvetlen működés során nem látható összefüggéseket.

A kutatási módszerek fontosságát nem mutatja markánsabban más, mint elterjedésük úgy az államigazgatás, mind az üzleti élet döntéseinek támogatásában. A legtöbb fejlett országban mind az állami szervek, mind a magánvállalatok rendkívüli erőfeszítéseket tesznek arra, hogy a kutatást támogató rendszereket kialakítsák, fejlesszék és a lehető leghatékonyabban működtessék. Ma már a világ több országában az állami programok sikeres és sikertelen előkészítésének, illetve utólagos értékelésének dokumentációja érhető el nagy tömegben. Ezekből világosan kitűnik, hogy a megfelelő ismeretek hiánya vagy birtoklása milyen drámai hatással lehet egy-egy ilyen program kimenetelére. Hasonlóan világos az is, hogy nemcsak a kutatási céllal használható minden adat, adatbázis fontos eleme a nemzeti adatvagyonnak, de mindaz a potenciális ismeret is, ami abból kinyerhető.

Az adatok hasznainak megítélésekor kiemelt figyelemmel kell lenni az ilyen, kutatás által indirekt módon előálló hasznok keletkezésére. A közvetlen adatfelhasználás esetében ez a legtöbb esetben nyilvánvaló: egy adott művelet sikere áll vagy bukik egy információ meglétén, és ennek a haszonnak az elismerése visszaköszön az élet legtöbb területén. Noha az adatforgalomban keletkező hasznok létezése kézenfekvő, a jelenlegi magyar jogi szabályozás és adatkezelési gyakorlat nem ismeri fel, nem mérlegeli az indirekt ismeretszerzés támogatása által a „köz” számára elérhető hasznokat, illetve azt a veszteséget, ami az ilyen felhasználás akadályozása által keletkezik.

4.6. Az adatok elemzésünk szempontjából releváns csoportosítása

Elemzésünk során a felhasználás módja, az adatból levonható következtetés általánossága szerint két adattípust különböztetünk meg: klasszikus és a kutatási célú közérdekű adatokat. Az elhatárolást elsősorban az alapján tesszük meg, hogy

az adott adattípust milyen következtetés levonására lehet használni. A klasszikus közérdekű adatra jellemzően egy döntési probléma, vagy problémák közvetlen megoldása érdekében van szükség, míg a kutatási célúnak nevezett kategóriába olyan adatok tartoznak, melyek felhasználási területe jellemzően az általános következtetések levonása. A történeti fejlődés tükrében a klasszikus adatok biztosítása sokkal inkább az állam szolgáltató funkciójához, míg a kutatói adatokhoz való hozzáférés inkább a kormányzati döntések meghozatalához kapcsolódik.

A klasszikus közérdekű adatok potenciálisan minden állampolgár érdeklődésére számot tarthatnak, és jellemző módon úgy hasznosulnak, hogy az adatkérő egy adott döntési helyzetében csökkentik a döntést övező bizonytalanságot. Ekkor olyan egyedi adatfelhasználásokról van szó, amelyek esetében az adatigénylő jellemzően nem természetes személyre vonatkozó információra kíváncsi. Az ilyen adatok elérhetővé tétele esetében általában az információ szolgáltatásának költsége dominál, hiszen létező adatokat kell elérhetővé tenni potenciálisan sok felhasználó számára, és igen kevés esetben jelentkeznek a személyes adatok védelmével kapcsolatos költségek.

A kutatói adatok esetében az egyedi esetre vonatkozó információ önmagában értéktelen. Az adatok annál hasznosabbak, minél „többben vannak”, minél nagyobb adatbázis áll rendelkezésre belőlük.¹³⁵ Elemzésükkel ekkor lehetséges egy tendenciát pontosan leírni, és ebből általános következtetéseket levonni. A kutatási célú adatok lehetnek személyesek, személyes jellegűektől azonban gyakran a kutatási érdek sérelme nélkül meg lehet fosztani azokat. A kutató számára ugyanis nem a személyességük fontos, hanem az egyediségük. Bizonyos alkalmazások esetében a személyesség költségmentesen megszüntethető, más alkalmazások esetén viszont csökkenhet az adat információtartalma, ha abból töröljük a személyes információkat. Az ilyen esetben az adatvédelem és a hasznosulás között átváltás van: minél többet megőrizünk az egyediségből, annál értékesebb lesz az adatok felhasználásával származtatott ismeret, ugyanakkor annál nagyobb az

¹³⁵ Leggyakrabban így van, de nem feltétlenül. Sokszor gond a haszontalan adattömeg, amelyből csak nagy munkával „bányászhatók ki” a másodlagos felhasználás céljára hasznos adatok. Ennek oka az, hogy az adatállományok nem a másodlagos felhasználásra készülnek.

esélye annak, hogy az adatokat felhasználva azonosítható annak alanya. Az adatok elzárása, a felhasználás tiltása a probléma megoldásának szélsőséges módja, amely nulla közvetlen költségen nulla közvetlen hasznot generál. Meg kell jegyeznünk, hogy a két adatforma szétválasztása nem tökéletes. Vannak olyan adatok, amelyek használatuk függvényében mindkét kategóriába besorolhatók. A meteorológiai adatszolgáltatás esetében például a potenciálisan minden állampolgárt érdekelhető heti hőmérséklet-előrejelzés klasszikus közérdekű adat, míg a saját időjárás-jelentésüket elkészítő információszolgáltatók által igényelt részletes műholdképek inkább kutatói adatnak minősíthetők. Besorolásunk tehát inkább *jellemző* elhatárolás, ami a későbbi kifejtésben igen hasznosnak bizonyul. A két adattípus főbb jellemzőit az alábbi táblázatban foglaljuk össze.

	Klasszikus	Kutatói célú
Felhasználás célja	Egyéni döntések hatékonyságának növelése	Általános következtetés levonása, döntés-előkészítés
Közvetlen felhasználás lehetősége	Igen, azonnali döntés hozható	Nem, elemző feldolgozás szükséges
Potenciális érdeklődők	Mindenki	Jellemzően kutatók
Kapcsolódó főbb költségek	Nyilvánosságra hozás	Adatelőkészítés Adatvédelem

4.7. A tanulmány szerkezete

Tanulmányunk közgazdasági eszközökkel tekinti át a közérdekű adatok szolgáltatásának mechanizmusát. A közérdekű, vagy azzá tehető adatok különböző felhasználási területein rámutatunk a belőlük származó hasznokra, amelyeket összevetünk előállításuk költségeivel. Ezzel párhuzamosan azt is vizsgáljuk, hogy a létező adatszolgáltatási intézmények gyakorlata mennyire van összhangban a közgazdasági következtetésekkel, egy olyan helyzettel, amelyben az adatok hozzáférhetőek. A gyakorlati tapasztaltok áttekintését nemcsak azért tartottuk fontosnak, hogy az elemzésünk tükrében ajánlásokat adhassunk az adatszolgáltatási mechanizmus esetleges javítására, hanem hogy rávilágítson

azokra a pontokra, amelyek elméleti elemzésünkéből egyébként kimaradnának. Ilyen mindenképp előtte az intézményi környezet és az ösztönzési rendszer, amiben az adatokat kezelő szervezetek működnek. Utólag azt mondhatjuk, a normatív és a leíró elemzésnek valóban kölcsönösen előnyös hatásai voltak egymásra.

A tanulmány második fejezete összefoglalja az elemzés szempontjából kulcsfontosságú közgazdasági eszközöket, és az azokra épülő következtetéseket. Ebben a részben a tisztánlátás érdekében csak elméleti síkon alapozzuk meg az adatok nyilvánosságára hozatalára és árazására vonatkozó elemzést. Eredményeink arra mutatnak, hogy az adatok nyilvánossága mindenképpen kívánatos. A jelenlegi törvényi szabályozással némileg ellentmondó következtetés ugyanakkor az, hogy a rendelkezésre álló források szűkössége miatt az adatokat a felhasználók egy részének érdemes térítés ellenében adni, amely adott esetben jelentős is lehet.

Ha ennyire egyértelmű a közérdekű adatok nyilvánossága, vajon megvalósul-e az a gyakorlatban? Előfordul-e, hogy névleges nyilvánosság mellett valójában nem hozzáférhetőek fontos adatok? Használják-e a potenciális felhasználók az adatforrásokat? Milyen pénzbeli és nem pénzbeli költségekkel jár a hozzáférés? A harmadik fejezet esettanulmányai ezeket a kérdéseket válaszolják meg azáltal, hogy a klasszikus közérdekű adatokhoz való hozzáférés magyarországi tapasztalatait tekintik át. Mivel teljeskörűsége törekednünk hiábavaló, olyan példákat igyekeztünk keresni, amelyek egyfelől a lakosság, és a vállalkozások számára releváns adatforrások, másfelől az adatkezelők stratégiáinak kellően széles skáláját lefedik.

A klasszikus és a kutatási célú adatok hozzáférhetősége közötti jelentős különbség, hogy az utóbbiak hozzáférését a jogi környezet nem támogatja kimondottan. Érzésünk szerint ennek fontos oka, hogy sem a jogalkotó, sem az adatgazdák számára nem teljesen világos a kutatási célú adatok haszna, hasznosulásának módja. Tekintve viszont, hogy itt túlnyomó részben egyéni, eredeti formájukban személyes adatokról van szó, felhasználásuk kockázatai annál nyilvánvalóbbak. A negyedik fejezetben bemutatjuk a kutatási adatokból realizálható hasznokat, és

példákon keresztül érvelünk amellett, hogy miért fontos az adatok elemi formájának nyilvánossága. A hasznok mellett részletesen foglalkozunk a nyilvánosságra hozatal során felmerülő költségeinek elemzésével is.

A kutatási adatok nyilvánosságának speciális, a klasszikus adatok esetében nem tapasztalt sajátága az adatvédelem, és a nem szakszerű értelmezés korábban csak részben érintett kérdése. Ennek megfelelően számos megoldás létezhet az adatok nyilvánosságára. Az ötödik fejezet a nemzetközi gyakorlat tapasztalatokat összegzi e téren. Itt bemutatjuk és összehasonlítjuk a jellemzőbb statisztikai adatszolgáltatási modelleket (Egyesült Államok, Egyesült Királyság, skandináv országok), valamint egyes nemzetközi szervezetek (IMF, Európai Unió) ajánlásai is. Ezek az áttekintések követendő példákat is állíthatnak, feltételezve, hogy a vizsgált országokban az adatgyűjtő és szolgáltató rendszerek kialakítói hasonló dilemmákkal álltak szemben, mint amit a negyedik fejezetben körülírtunk, és ezekre működőképes megoldásokat találtak.

Amint azt a klasszikus közérdekű adatok tárgyalásakor is hangsúlyoztuk, az egyes adatgazdák adatkezelési szokásai és érdekei alapvetően befolyásolhatják, hogyan viszonyulnak a nyilvánossághoz. A hatodik fejezet áttekinti azon hazai intézmények működését, amelyek már ma is szolgáltatnak mikroadatokat, vagy erre a kezelésükben levő adatvagyon birtokában képesek lennének.

A tanulmány utolsó, hetedik fejezete először egy gondolkísérletet mutat be. Ebben feltesszük, hogy az Avtv. rendelkezései teljes mértékben érvényesülnek: minden közérdekű adat a közzététel költségétől eltekintve ingyenesen elérhetővé válik, majd megvizsgáljuk, várhatóan hogyan változna az adatok kereslete, illetve milyen további költségekkel kell számolnunk a kínálati oldalon. A kísérlet alapján ajánlásokat fogalmazunk meg arra vonatkozólag, hogyan érvényesíthető hatékonyan az információs szabadság elve.

5. Az adatszolgáltatásra vonatkozó elméleti eredmények

Ebben a fejezetben áttekintjük közgazdasági elemzésünk folyamán használt alapvető fogalmakat, valamint ezek alapján megfogalmazzuk azokat az elméleti eredményeket, amelyekre a későbbi elemzésekben visszautalunk. Első eredményünk annak megmutatása, hogy finanszírozási korlát nélkül társadalmilag optimális a közérdekű adatok hozzáférhetőségének ingyenes biztosítása. Fő állításunk az, hogy az adatszolgáltatás árának csak az információátadás költségét kell tartalmaznia, amely összhangban van az Avtv. előírásaival is.

A folyamatos adatgyűjtés és az adatszolgáltatás infrastruktúrájának fenntartása ugyanakkor igen nagy költséggel jár, amely független az adatforgalomtól, és ehhez a forrásokat az államnak valamilyen módon biztosítani kell. Amennyiben lehetőség van arra, hogy közvetlen költségvetési juttatásokból (vagyis adóbevételekből) fedezzék ezeket a költségeket, akkor hatékony biztosítani az ingyenes hozzáférhetőséget. Ha viszont a költségvetés szűkössége ezt nem teszi lehetővé, akkor érdemes lehetőséget biztosítani az adatszolgáltató szervek számára, hogy az adatok értékesítésével piaci úton szerzett bevételekből ellensúlyozzák a rendszer működtetéséből származó veszteségeiket. Amennyiben erre az adatszolgáltatóknak nem adunk lehetőséget, az hosszú távon a szolgáltatás minőségének romlásához is vezethet.

A fejezet második részében bemutatjuk, hogy az árazás megengedése milyen el nem kerülhető hatékonyságveszteségeket okoz, illetve milyen módon csökkenthetőek ezek a torzító hatások. Bemutatjuk, hogy az optimálisan választott diszkriminatív árazás magasabb jólétet eredményez, mint ha egységes árakat szabunk meg az adatigénylők különböző csoportjainak. Végül szót ejtünk arról az adatszolgáltatás és feldolgozás folyamán felmerülő externális hatásokról, és ezek figyelembevételéről az optimális árazási politikánál.

Érdemes megjegyeznünk, hogy bár a tárgyalásmódunk eredendően az optimális árak szempontjából közelíti meg a vizsgált kérdést, ez ekvivalens a hozzáférés kérdésének a vizsgálatával. Ha ugyanis az adatszolgáltatás ára nulla, akkor ez szabad hozzáférést jelent az adatkérők számára, az ár növekedésével pedig a biztosított hozzáférés folyamatosan csökken.

5.1. Az adatszolgáltatás közjóság jellege

Egy szolgáltatást akkor nevezünk *közjóságnak*, ha egy gazdasági szereplőnek történő biztosítása esetén a további gazdasági szereplők is költségmentesen hozzájuthatnak ehhez a szolgáltatáshoz. A közjóság egyik specialitása, hogy az azt használók között *nincs rivalizálás*, vagyis egy szereplő „fogyasztása” következtében a többi szereplő által igénybe vehető mennyiség nem csökken.¹³⁶ Ez a tulajdonság adatok esetben nyilvánvalóan teljesül. A közjóságok másik specialitása, hogy fogyasztásukból *nem zárhatók ki* egyes szereplők. Ennek a definíciónak az értelmében a nyilvánosan közzétett adatok a közjóságok tipikus példái, amilyen az internetes honlapról letölthető adatbázis vagy egy önkormányzat ügyfélszolgálati rendjét hirdető tábla.

Az adatszolgáltatás önmagában azonban még nem tekinthető közjóságnak, ugyanis az adat kezelőjének lehetősége nyílhat árazási vagy diszkriminatív eszközökkel egyes szereplőket kizárni a hozzáférésből, vagyis az ő tudatos döntése, hogy az adat mennyire ölt közjóság-jelleget.¹³⁷ Közgazdasági következtetéseinket tehát nem ebből a tulajdonságból kell levezetnünk, mivel ez a tulajdonság már következmény. Ezért is koncentrálunk az olyan háttérben álló okokra, mint az adatszolgáltatás költségstruktúrája, illetve az adatok „fogyasztásából” származó pozitív externális hatások.

5.2. Hasznok és költségek, társadalmilag optimális ár

¹³⁶ Rivalizálás merül fel például egy 10 férőhelyes parkoló esetében, amely ugyan ingyenesen igénybevehető, de csak az első 10 odaérkező gépkocsi számára.

¹³⁷ A klasszikusan közjóságnak tekintett javak, mint például a nemzetvédelem vagy a közvilágítás esetén ez a kizárhatóság nem lehetséges.

Tekintsük át először a hasznok és költségek összevetésének alapvető módszerét, valamint hogy ennek az egyszerű, közgazdaságtanban általánosan használt modellkeretnek a segítségével mit mondhatunk árazási kérdésekről!

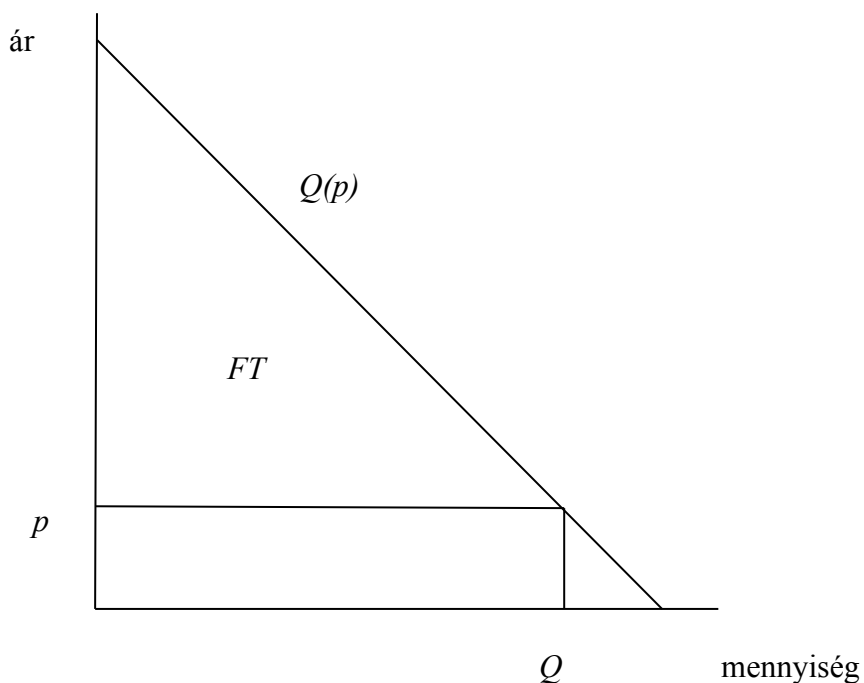
Kérdésünket egyszerűsítsük le arra a kérdésre, hogy egy állami szervezetnek már birtokában van egy bizonyos adat, ami potenciálisan minden állampolgárt érdekelhet, de különböző mértékben. Vegyük például a családjogi esetekben hozott bírósági döntések kivonatos gyűjteményét az elmúlt 10 évre nézve. Egyes professzionális felhasználók számára, amilyen egy ilyen ügyekre specializálódott ügyvédi iroda, ez az adatbázis igen nagy értéket képviselhet. Használatával jelentősen csökkentheti felkészülési költségeit, míg egy gazdasági ügyekkel foglalkozó iroda számára kisebb hasznot jelent az adatbázishoz való hozzáférés. Egy ilyen adatbázis különböző mértékben érdekelhet egyes állampolgárokat is, ami nyilván annak függvénye, hogy családi állapotuk mennyire teszi azt szükségessé.

Azt, hogy az egyes felhasználók mennyire értékelnek egy adott terméket, a mi esetünkben az adatot, a következőképpen vethetjük össze: feltesszük nekik (hipotetikusan) azt a kérdést, maximálisan mekkora összeget lennének hajlandók adni az adott termékért, vagyis mennyi pénzt érne meg nekik a hozzáférés. Az egyes potenciális fogyasztók által megnevezett árat az adott személyek *rezervációs árainak* nevezzük. Amennyiben ismerjük minden személy rezervációs árát, könnyen meghatározhatjuk a jószág iránti *keresletet*: bármely áron az adatnak annyi fogyasztója (igénylője) lenne, amennyinek a rezervációs ára magasabb, mint a tényleges piaci ár. Könnyű belátni, hogy ahogy a piaci ár csökken, a jószág által támasztott kereslet nő.

A következő kérdés az, hogy amennyiben az adathoz való hozzáférést egy adott áron biztosítják, hogyan számszerűsíthető ezen ár mellett a fogyasztók összhaszna, más néven jóléte. Amennyiben például egy fogyasztó az adathoz való hozzáférésért 100 ezer forintot is megadna, de az ára, amit ténylegesen fizetnie kell érte, csak 40 ezer forint, akkor neki (nettó) 60 ezer forint haszna származik az adat megszerzéséből. Amennyiben ezeket az egyéni szinten jelentkező többlethasznokat

összesítjük az összes vevőre nézve, akkor megkapjuk az úgynevezett (*nettó*) *fogyasztói többletet*.

Az eddig áttekintett fogalmak szemléltetésére tekintsük az 1. ábrát. Tegyük fel, hogy a potenciális fogyasztók száma kellően nagy ahhoz, hogy a keresletet leírhatjuk egy folytonosan csökkenő görbével, és jelölje az adott p ár melletti keresletet $Q(p)$.¹³⁸ Amennyiben a keresleti görbét a mennyiség felől tekintjük, akkor például annak a $Q = 20$ mennyiségnél felvett értéke megmutatja a 20. legnagyobb rezervációs árral rendelkező fogyasztó rezervációs árát. Mivel az egyes fogyasztók többlethaszná egyedi rezervációs árak és a tényleges p ár különbsége, ezért a teljes fogyasztói többlet (az ábrában FT) a keresleti görbe és a p árnál húzott vízszintes egyenes közötti terület nagysága. Amennyiben az adathoz való hozzáférést ingyenesen biztosítják, akkor a fogyasztói többlet megegyezik a keresleti görbe alatti teljes terület nagyságával.

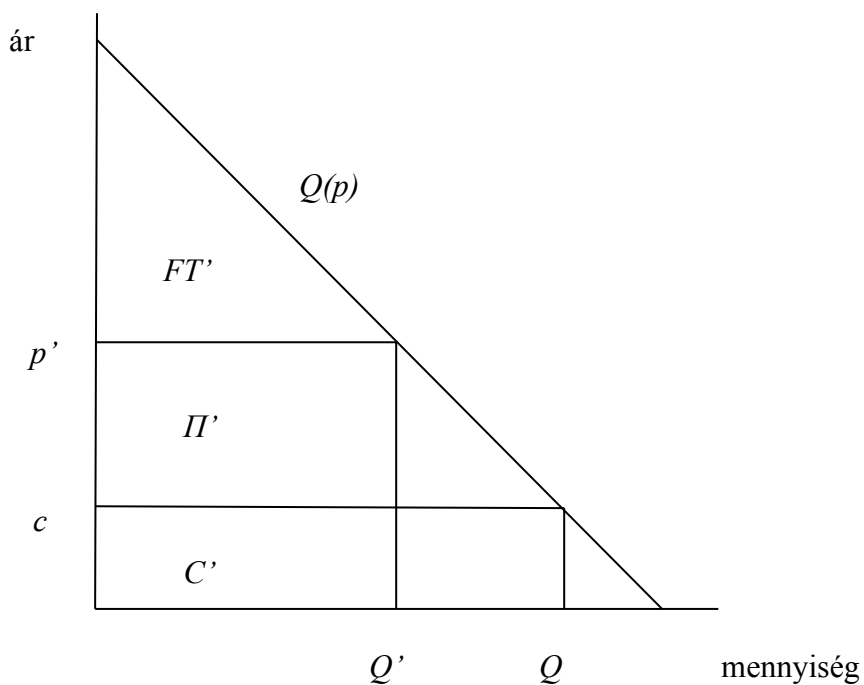


¹³⁸ Bár a piacon keresett mennyiség az ár függvénye, a közgazdaságtanban elterjedt 100 éves konvenció szerint a p árat (mint független változót) ábrázoljuk a függőleges, és a Q mennyiséget (mint függő változót) a vízszintes tengelyen.

1. ábra

Az adott áron való adathozzáférés hasznainak áttekintése után vizsgáljuk meg a másik oldalt, vagyis hogy hogyan mérhetjük az adatszolgáltatás költségeit. Az adat, mint úgynevezett információs jószág, termelési folyamata speciális költségstruktúrával rendelkezik: az első példány előállítása nagy költségekkel jár, de a további példányok már igen kis költséggel reprodukálhatóak. Közgazdasági szaknyelven azt mondhatjuk, hogy a termelés *fix költsége* magas (jelölje ezt F), de a pótlólagos egységek előállítási költsége, az úgynevezett *határköltség* alacsony és gyakorlatilag konstansnak tekinthető (jelölje ezt c).

Gyakran megjelenő állítás, amit a következő alfejezetben még kicsit alaposabban megvizsgálunk, hogy a vizsgálandó adatszolgáltató szervezeteknek vagy törvényi kötelessége a kérdéses adatok gyűjtése, vagy tevékenységük melléktermékeként állnak ezek elő, tehát az F összeget mindenképpen meg kell fizetni. Amikor tehát ahhoz a kérdéshez érkezünk, hogy milyen széles körben (azaz milyen áron) kell biztosítani egy adott adathoz a hozzáférést, a fix költség már kifizetett, úgynevezett *elsüllyedt költség*. Ebben az esetben az optimális hozzáférés mértékének meghatározásánál figyelembe veendő költségek tehát csak a termelés úgynevezett változó költségei, ami tetszőleges Q mennyiség esetén $c \cdot Q$.



2. ábra

Összegezzük most a hasznokat és költségeket egy tetszőleges $p' \geq c$ ár mellett, amelyeket a 2. ábrán szemléltetünk. A p' ár melletti kereslet Q' , ekkor a fogyasztói többlet FT' . A Q' mennyiség szolgáltatásának változó költsége $c \cdot Q'$ (ezt az összeget jelöli a C' terület), az adatszolgáltató bevétele pedig $p' \cdot Q'$, vagyis profitja $(p' - c) \cdot Q'$, amit az ábrán Π' jelöl.¹³⁹ Amennyiben a teljes társadalom jólétébe beleszámoljuk az adatszolgáltató bevételeit is (mondjuk olyan érvelés alapján, hogy a finanszírozás terheit így nem az adófizetők viselik, vagy ezzel mérjük a munkahelyek által teremtett jólétnövekedést), akkor p' ár mellett kialakuló

¹³⁹ Ebből a profitból még le kell vonnunk a fix költséget, amit az ábrán nem tudtunk ábrázolni.

társadalmi többlet (vagy más néven jólét) nagysága $FT' + \Pi'$, vagyis a keresleti görbe és c határköltségnél húzott vízszintes egyenes közötti terület nagysága a Q' mennyiségig.

Az elmondottak alapján most már egyszerű megválaszolnunk azt a kérdést, hogy mi legyen az adatszolgáltatás társadalmilag optimális, azaz jólét-maximalizáló ára. Amennyiben $p' > c$, akkor az ár kis csökkentésével, azaz a hozzáférés kis mértékű megnövelésével növekszik a társadalmi többlet, hiszen az alacsonyabb ár mellett bekapcsolódó fogyasztók hasznossága nagyobb, mint a pótlólagos hozzáférés megteremtésének határköltsége. A határköltségnél kisebb árat pedig nem érdemes választanunk, hiszen ekkor az újabb fogyasztók bekapcsolódásának társadalmi határhaszna kisebb, mint annak határköltsége. A társadalmilag optimális ár tehát egyenlő az adatszolgáltatás határköltségével, függetlenül a kereslet sajátosságaitól.

Mivel az adatszolgáltatás határköltsége közel nullának tekinthető, ez az állítás egyenértékű azzal, hogy társadalmilag optimális az állam kezelésében lévő adatvagyon teljes nyilvánossága. Az Avtv. is pontosan ezt írja elő, amikor azt mondja, közérdekű adatoknál költségtérítés a legfeljebb a közléssel kapcsolatban felmerült költség mértékéig terjedhet, a tényleges költségek megtérítésén túl profitelemet nem tartalmazhat.

Fontos megjegyeznünk, hogy bár úgy érveltünk, mintha ismernénk a fogyasztók adott termékre vonatkozó rezervációs árait, ennek az alapvető eredménynek az eléréséhez nincs erre szükségünk; mindössze a határköltség nagyságát kell tudnunk, és azt, hogy a keresleti görbe negatív meredekségű.

Megfigyelhetünk-e határköltség alapú árazást a gyakorlatban? Sokat idézett példa a teljes nyilvánosságra az Interneten közzétett információk és adatbázisok. A szolgáltató, akinek már rendelkezésére áll az adatbázis és a nyilvánosságra hozás technikája (a honlap), gyakorlatilag elhanyagolható költséggel kiadhat egy újabb információt, és annak fogyasztói ezért nem fizetnek semmilyen árat. Meg kell jegyeznünk, hogy ettől még korántsem biztos, hogy a fogyasztók oldalán nem

merülnek fel további költségek (például az Internet-előfizetés fix költsége vagy a keresés határköltsége), de pénzbeli kifizetés a szolgáltató felé nem történik.

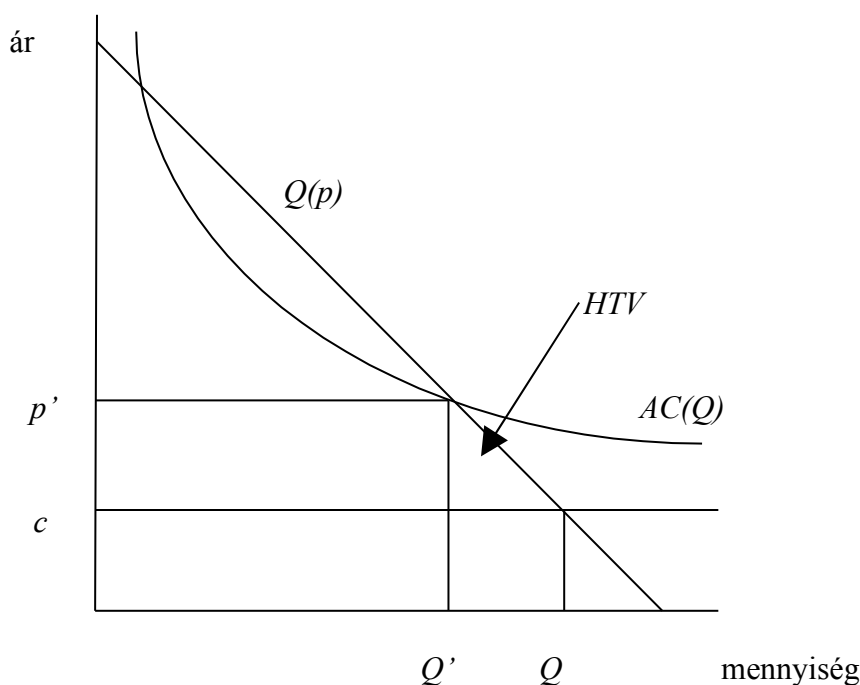
A társadalmilag optimális határköltség-alapú árazás alapgondolata tehát az, hogy a piacon egyetlen tranzakció se vesszen el, amelyik értékes abban az értelemben, hogy annak haszna meghaladja annak költségét. A tranzakciók megvalósulásához szükséges speciális beruházás (jelen esetben az adatbázis előállítás és fenntartása) költségigényének fedezése ugyanakkor nem jelenik meg ebben a gondolatmenetben. Alkalmazásakor lényegében azzal a feltevéssel élünk, hogy amennyiben a hasznos tranzakciók összesített nettó haszna (a nettó fogyasztói többletek összege) meghaladja a beruházás fix költségét, akkor társadalmilag optimális ezt a beruházást megtenni, de hogy ennek terheit ki viselje, arról nincs szó.

5.3. Az adatszolgáltató szervezetek bevételigénye és ennek árazási következményei

Az előzőekben levezetett társadalmilag optimális (úgynevezett első legjobb) határköltség-alapú árazás gyakorta felvetett problémája, hogy alkalmazása mellett a szolgáltató profitja negatív. Ennek oka a speciális költségszerkezetben keresendő: a konstans határköltség és a magas fixköltség jelenléte azt eredményezi, hogy a termelés egységnyi mennyiségre jutó úgynevezett átlagköltség (average cost) mutatója, vagyis az $AC = (F + cQ) / Q = F/Q + c$ érték csökkenő, valamint mindig a határköltség c szintje felett marad, miképpen ez a 3. ábrában is látható.

A határköltség-alapon választott $p = c$ ár mellett a szolgáltató bevételei így csak a $c \cdot Q$ változó költségeket fedezik, tehát F veszteséggel zár, amit az állami költségvetésnek kell finanszírozni. Amennyiben felmerül az igény, hogy az adatszolgáltató teremtsen meg a működési költségeit fedező bevételeket (ezt röviden profitabilitásnak nevezzük), milyen árat engedélyezzünk számára, amely ezen cél mellett maximalizálja a társadalmi többletet? Mivel a társadalmi többlet az ár csökkenő függvénye, a szolgáltató profitabilitása pedig akkor teljesül, ha az ár (ami

egyben az egységnyi mennyiségre jutó átlagos bevétel) nem kisebb, mint az átlagköltség, ezért azt a p' árat érdemes választani, ahol az átlagköltség-görbe az alacsonyabb árnál elmetshi a keresleti görbét.



3. ábra

Fontos észrevennünk, hogy az átlagköltség-alapú árazás kisebb jólétet eredményez, mint a határköltség-alapú, ezért ezt a kimenetet szokás „második legjobb optimumnak” is nevezni. Mindez azon az egyszerű következményen alapul, hogy átlagköltség-alapú árazás mellett a Q' és Q közötti fogyasztók nem részesülnek a szolgáltatásból, mert az rontaná a szolgáltató profitabilitását. Ezeknek a fogyasztóknak ugyanakkor társadalmilag optimális lenne szolgáltatni, mivel az

alacsonyabb ár mellett bekapcsolódó fogyasztók hasznossága nagyobb lenne, mint a pótlólagos hozzáférés megteremtésének határköltsége. Ezt a Q' és Q között elvesző társadalmi többletet szokás holtteher-veszteségnek is nevezni, ezt az ábrán *HTV* jelöli. Ha tehát a szolgáltatásért csak egy egységárat kérünk, akkor az adatszolgáltató profitabilitásának biztosítása és a jólétet maximalizáló határköltség-alapú árazás együttesen nem megvalósítható.

Mint látni fogjuk, ez az adatszolgáltatás-árazási elv, miszerint az adatokat „önköltségen” kell adni, számos adatszolgáltató szervezet árazási struktúrájában megtalálható. Első ránézésre furcsának tűnhet, hogy a bevételszerzést egy állami szerv feladatául szabjuk. Ugyanakkor vegyük észre, hogy egy szűkösségi problémával állunk szemben, a költségvetés ugyanis csak korlátozott erőforrásokat képes egy-egy feladathoz rendelni, ami a legtöbb esetben nem elég a társadalmilag optimális beruházásokhoz. Alternatív megoldásként felmerülhetne az, hogy az állam megemeli az adókat annak érdekében, hogy a pótlólagosan befolyt összegből végrehajtja a társadalmilag optimálisnak nyilvánított határköltség-alapú árazás tényleges megvalósulását. Ugyanakkor amennyiben bizonyos állampolgárok nem veszik igénybe az adatszolgáltatást, mert pótlólagos hasznaik a mégoly alacsony határköltség alatt vannak, akkor nem csak az adatszolgáltatás haszonélvezőivel fizettetjük meg a szolgáltatás létrejöttét. Ráadásul arra egyetlen költségvetésben sincs garancia, hogy a befolyt bevételeket a megígért célokra költik, ezzel szemben az adatszolgáltató szervezethez befolyó bevétel lehetővé teszi, hogy azt saját költségeinek fedezésére fordítsa.

Bár adott pillanatban elmondható, hogy a már meglévő adatok gyűjtésének költségét már kifizették, tehát ezeket az éppen aktuális árazási és hozzáférési politikánál nem kell figyelembe vennünk, felmerülnek további költségek, amelyeket nem szabad figyelmen kívül hagynunk. Az első a meghatározott adathozzáférési szint biztosításának a költsége, ami annál magasabb, minél alacsonyabb az ár (vagyis minél nagyobb a kereslet). Másrészt a hozzáférés színvonalát biztosítani kell, ami újabb adatok felvételéből, az adatbázisok aktualizálásából áll, és ezekre a jövőben jelentkező költségekre meg kell teremteni a forrást. Amennyiben ezt a

költségvetés nem képes biztosítani, akkor érdemes megengedni, hogy az adatszolgáltató szervek *szabályozott árakból* származó bevételeikből ellensúlyozzák a rendszer működtetéséből származó veszteségeiket. Amennyiben erre nem adunk lehetőséget, hosszú távon ez a szolgáltatás minőségének romlásához vezethet.

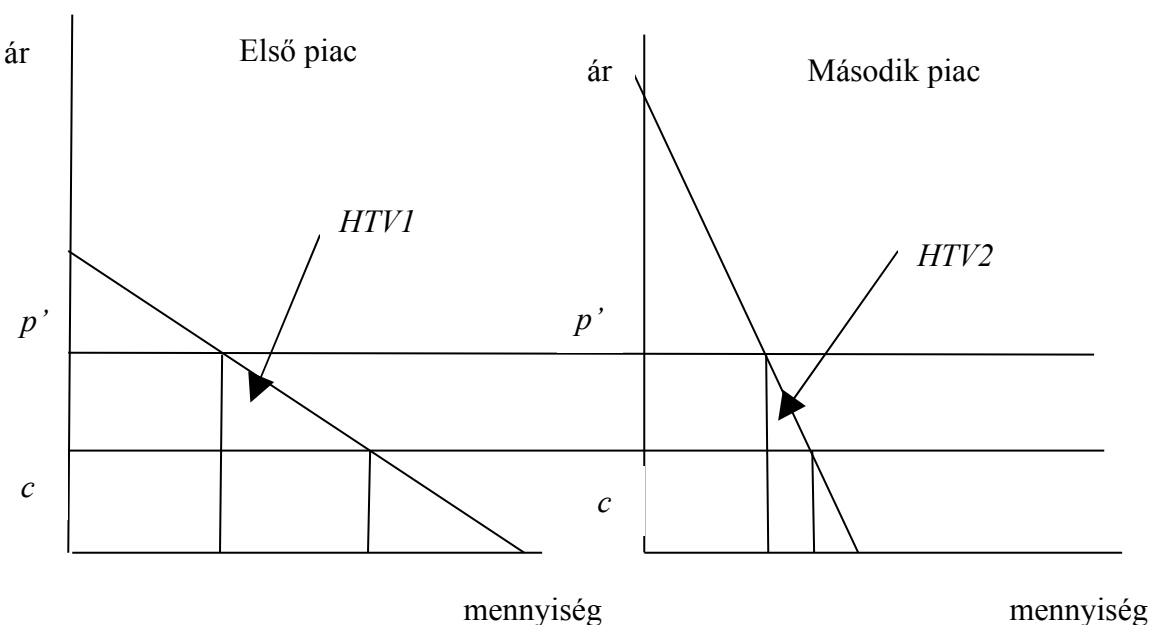
Ha az állam rákényszeríti ezt az alapelvet intézményeire, azzal lényegében „saját zsebéből húzza ki a pénzt”, még ha nem is konkrét adóforintok, de nem realizált társadalmi hasznok képében. Ugyanakkor ha ezt nem teszi meg, az adatszolgáltató szervezet veszteségessége hosszú távon veszélyeztetni annak működését, ami nagyobb hatékonyságvesztést okozhat.

Bár ezt a megközelítést nem elemezzük részletesen, a minimum „nullszaldósság” elvárása mögött néha az a gondolat is megjelenik, hogy a piaci viselkedés költséghatékonyabb belső működés kialakítására ösztönzi az adatszolgáltatót. Mindez igaz lenne, amennyiben a költségmegtakarításokról döntő vezetők jövedelme attól függene, mennyire nyereséges az adott adatszolgáltató, amitől Magyarországon ma még messze állunk. Sokkal inkább érvényesül a „puha költségvetési korlát” hatása, miszerint mindenki előre látja, hogy veszteségeit az államkassza majd betömi, így nem érdemes egyéni erőfeszítéseket tenni a költségek csökkentése érdekében.

5.3.1. Az adatszolgáltató bevételeigénye és árdiszkrimináció

Az előző alfejezetben láttuk, hogyha az adatszolgáltató egységes árat állapít meg minden vevője számára, és biztosítani kell a profitabilitását, akkor átlagköltség-alapú árazást kell alkalmaznunk. Felmerülhet azonban az a kérdés, hogyha a szolgáltatott jószágot két vagy több eltérő jellemzővel bíró felhasználói csoport alkalmazza, akkor optimális-e számukra különböző árakat, azaz árdiszkriminációt alkalmazni?

Meglátásunk szerint közérdekű adatok szolgáltatása esetén igen jól elkülöníthető két ilyen csoport, az egyszerűség kedvéért nevezzük őket állampolgároknak és piaci fogyasztóknak.¹⁴⁰ Az első fogyasztói csoportba sorolt fogyasztók száma igen nagy, de egyéni keresletük jellemzően igen kis mennyiségű; kevésbé érzékenyek az adatok minőségére, de annál inkább az árra. Ezzel szemben a második fogyasztói csoportba eső fogyasztók száma kisebb, viszont egyenként sokkal nagyobb mennyiségű adatra tartanak igényt; az árra relatíve kevésbé, a szolgáltatás minőségére viszont annál érzékenyebbek. Közgazdasági fogalmakkal élve azt mondhatjuk, hogy adott ár mellett az első piac *árrugalmassága* nagyobb: az ár egy százalékos emelkedése az első részpiacon sokkal nagyobb relatív (azaz szintén százalékban mért) keresletcsökkenéssel jár, mint a második részpiacon.¹⁴¹



¹⁴⁰ Ha csak a professzionális felhasználók által keresett kutatói adatokról beszélünk, akkor az első csoportba a tudományos kutatókat, a második csoportba a piaci felhasználókat sorolhatjuk.

¹⁴¹ Professzionális felhasználók azért is kevésbé érzékenyek az árra, mivel nekik tevékenységük folytatásához általában nélkülözhetetlenebb az adott adat. Számos empirikus kutatás bizonyítja, hogy az árrugalmasság annál kisebb, minél nagyobb az adott szereplő jövedelme, részben ez is indokolja a két csoport árrugalmassági alapon történő elkülönítését.

4. ábra

A két eltérő tulajdonsággal rendelkező keresletre láthatunk példát a 4. ábrán. Látható, hogy amennyiben a c határköltségnél magasabb p egységes árat alkalmazzuk, akkor az első (állampolgári) részpiacon a határköltségtől eltérő ár sokkal nagyobb holtteher-veszteséget okoz, mint a második piacon ($HTV1 > HTV2$). Ahhoz, hogy a két piacon elért hatékonyságveszteség összegét minimalizáljuk, érdemes az első részpiacon alacsonyabb árat választani, mint a második részpiacon. A közgazdaságtan ezt az árazási formát Ramsey-árazásként ismeri és ennek a megközelítésnek a fő eredménye szerint az első és második részpiacon alkalmazott optimális p_1 és p_2 árakat úgy kell megállapítani, hogy

$$\frac{p_1 - c}{p_1} \varepsilon_1 = \frac{p_2 - c}{p_2} \varepsilon_2$$

legyen, ahol ε_1 és ε_2 az első és a második részpiac árrugalmasságát jelöli. A rugalmasságok azért ragadják meg jól ennek a problémának a lényegét, mert úgy szeretnénk növelni a határköltségnél magasabb szintre az árakat a profitabilitás eléréséhez, hogy az minél kevesebb kereslet-kieséssel járjon, és ezért azon a piacon kell nagyobb ártorzítást alkalmaznunk, ahol annak kisebb hatása van a termék keresletére.

Ezen eredmény szerint a magasabb árrugalmasságú piacon kisebb mértékben kell eltérni a határköltségtől. Amennyiben például az ár kellően közel van a majdnem nulla határköltséghez, az állampolgárok részpiacán az árrugalmasság értéke messze meghaladhatja az üzleti felhasználók piacának árrugalmasságát (számos állampolgár nulláról pozitívrá változtatná adatfogyasztását, ha az hirtelen ingyenessé válna), vagyis ennek alapján azt mondhatjuk, hogy az állampolgárok felé alkalmazott árak egyenlőnek kell lennie a határköltséggel, míg az üzleti felhasználók felé alkalmazott árból kell fedezni az adatszolgáltató fix költségeit. Újabb eredményünk tehát az, hogy amennyiben el kell térnünk a határköltség-alapú árazástól, az árdiszkrimináció engedélyezése növeli a jólétet.

Jegyezzük meg ugyanakkor, hogy az Avtv. egyértelműen kimondja, hogy az adatszolgáltatásért kért költségtérítés összege nem tehető függővé a közérdekű adat felhasználásának céljától, vagyis tiltja az ilyen fajta diszkriminációt. A törvény lényegében a határkölség-alapú árazást támogatja, és annak alkalmazása esetén a diszkrimináció valóban jólétsökkentő hatású lenne. Amennyiben viszont el kell térnünk a társadalmilag optimális árazástól, az egységes ár alkalmazása többé már nem hatékony, a diszkrimináció engedélyezésével növelhető a jólét.¹⁴²

Érdemes észben tartunk ugyanakkor, hogy ez a fajta árdiszkriminációs irány egybevág azzal, ami akkor következne be, ha az adatszolgáltatónak megengednénk, hogy profitját maximalizálná. Ebben az esetben is a kevésbé árérzékeny felhasználói csoport irányába szabna nagyobb árat, ugyanakkor tovább emelné az árat mindkét részpiacra. Pusztán abból a tényből tehát, hogy diszkriminatív árazást figyelünk meg és az árak a határkölségnél magasabbak, még nem tudjuk elkülöníteni, hogy az a második legjobb társadalmi optimumhoz, vagy egy profitmaximalizáló szolgáltatóhoz tartozik, ehhez meg kell tudnunk pontosan megfigyelni a szolgáltató bevételeit és összevetni azt kiadásaival. Amennyiben úgy látjuk, hogy az adott szolgáltató „túlzóan” profitorientált diszkriminatív árpolitikát folytat, arra következtethetünk, hogy a különböző fogyasztói csoportoknak szabott árakat arányos kell csökkenteni az indokoltnak tekintett profitszint eléréséig.

5.3.2. Ár- és minőségi diszkrimináció

Az előző részben bemutatott (úgynevezett harmadfokú) árdiszkriminációs megkülönböztetés azon alapul, hogy a szolgáltató képes elkülöníteni a két részpiac szereplőit, és azoknak különböző árakat megszabni. Ez a feltevés nem mindig teljesül, az alacsony ár reményében az üzleti felhasználóknak érdekük fűződhet ahhoz, hogy állampolgári jogukra hivatkozva kérjék az adatokhoz való hozzáférést. Hogyan biztosítható mégis az, hogy az üzleti felhasználóknak érdekükben álljon magasabb árat fizetni?

¹⁴² A közgazdaságtan ezt az állítást a „Második Legjobb Tételének” nevezi: egy hatékonyság-torzító lépés alkalmazása esetén még egy önmagában hatékonyság-torzító elem is jólétnövekedést eredményezhet.

Az árérzékenységen kívül általában létezik egy további jellegzetes eltérés a két fogyasztói csoport között: az üzleti felhasználók többre értékelik a minőséget. Ez a tulajdonság lehetővé teszi, hogy amennyiben az adatszolgáltató képes magasabb minőségű adatszolgáltatásra, azt olyan áron tegye elérhetővé, hogy az üzleti felhasználóknak ezt álljon érdekükben választani. Az elmondottak érzékeltetésére tekintsük a következő példát.

Tegyük fel, hogy a piacon egy állampolgár és egy cég van, akik egységnyi keresletet támasztanak a szóban forgó adat iránt, amely lehet kielégítő vagy jó minőségű. A cég rezervációs ára a két minőség iránt legyen 4 és 15, az állampolgáré 3 és 7, az adatszolgáltatás határkölsége legyen zérus (az egységeket vegyük ezer forintban).

Amennyiben a szolgáltató csak egy árat / minőséget választhat, akkor csak a céget érdemes kiszolgálni jó minőségű adattal, a szolgáltató profitja ekkor 15, a cég teljes hasznát viszont az adatszolgáltató bezsebeli.

Minőségi diszkrimináció esetén a kielégítő minőségű termék árát 3 egységben érdemes megállapítani (az állampolgár ezt fogja megvenni), míg a jó minőséget továbbra is az első fogyasztónak szánja, de annak árát le kell csökkentenie 14-re, mivel a kielégítő minőség megvételével a cég $4 - 3 = 1$ többletet kaphatna. A szolgáltató profitja ekkor megnő 17-re, a cég szigorúan jobban jár, az állampolgár helyzete nem változik. Amennyiben diszkrimináció esetén megköveteljük, hogy az adatszolgáltatás kielégítő (úgynevezett minimális) szintje ingyenes legyen, a szolgáltató akkor is elkérhet 11 egységnyi árat a jó minőségért, és azt a cégnek érdekében áll megvennie, és így mindegyik fogyasztó határozottan jobban jár a diszkrimináció bevezetésével.

Ez a példa ismét rámutat arra, hogy az ár- és minőségi diszkrimináció engedélyezése növelheti a jólétet. A jólétnövekedés szükséges feltétele, hogy a diszkrimináció engedélyezése növelje a részpiacokon együttesen támasztott

összkeresletet,¹⁴³ amely feltétel megítélésünk szerint az állampolgári – üzleti elkülönítésre épülő piacszegmentálás esetén könnyen teljesül.

Bár adatok esetén sokkal elterjedtebb a minőségi diszkrimináció, egyes piacokon előfordul, hogy úgynevezett mennyiségi diszkriminációt alkalmaznak. Ennek egyik jellemző formája lehet például, hogy két különböző nagyságú adatcsomagot hoznak létre, a több adatot tartalmazó csomag ára ugyan összességében magasabb, az adatok egységára ugyanakkor mégis alacsonyabb.¹⁴⁴ Ez a két csomag szintén alkalmas lehet a két felhasználói csoport elkülönítésére, amennyiben az egyik csoport tagjai egyenként kis keresletet támasztanak az adatok iránt (az állampolgárok többségét általában csak az őt közvetlenül érintő adatok érdeklik), míg a másik csoport tagjainak potenciálisan igen nagy az adatigénye (az üzleti felhasználók igazán az adatbázisoknak veszik hasznát).

A jobb minőségű adatokat ezzel lényegében a felhasználó személyéhez és / vagy céljaihoz szabják, így egyre inkább privát jószággá válnak, amelyeket igazán csak az adott felhasználó értékel. Ebben az értelemben még inkább érthető az a diszkriminatív árazás mögött húzódó alapgondolat, hogy ezeket a speciális igényű felhasználókat tekintsük bevételi forrásnak, amely részpiacon történő értékesítésből fedezhetőek az adatszolgáltatás fix költségeit. Ezzel szemben az alapszolgáltatás esetében a diszkrimináció nem érvényesíthető, ezért ott egységes (határköltséghez közeli) árazást érdemes alkalmazni minden szereplő felé.

5.3.3. Kétrészes tarifák alkalmazhatósága

Láttuk, hogy a határköltség alapú árazással szemben gyakran felmerülő kritika az, hogy az adatszolgáltatásra jellemző költségstruktúra mellett veszteséget eredményez az adatszolgáltatóknál. Ez az eredmény ugyanakkor szoros összefüggésben áll azzal a feltétellel, hogy szolgáltató számára csak azt engedjük meg, hogy a szolgáltatott mennyiségért egy egységes (darab)árat határozzon meg.

¹⁴³ Ezt a ma már klasszikusnak számító eredményt elsőként Schmalensee (1981) és Varian (1985) látták be.

¹⁴⁴ Ez a jelenség a szupermarketekből jól ismert diszkont esete.

A magasabb ár magasabb profithoz, ugyanakkor kisebb társadalmi többlethez vezet, tehát úgy tűnhet, mintha leküzdhetetlen ellentmondással néznénk szembe.

Ez a feladat azonban korántsem megoldhatatlan, ha nem egy eszközzel próbáljuk megoldani mindkét célt. Megengedhetjük azt, hogy az adatszolgáltató megszabhasson egy fix összeget, úgynevezett belépési vagy előfizetési díjat, és csak azoknak a személyeknek szolgáltat meghatározott egységaron, akik előfizettek a szolgáltatásra. Az egységarat meghatározhatjuk úgy, hogy az pontosan fedezze a változó költségeket, a belépési díjat pedig úgy, hogy azokból összességében megtérüljenek a fix költségek, vagyis ez a tarifarendszer lényegében a költségek szerkezetét másolja le. Ennek az optimálisan választott úgynevezett *kétrészes tarifa* két különböző eszköze alkalmas tehát arra, hogy külön-külön oldják meg a két feladatot: az egységarat egyenlővé tehetjük a határköltséggel, így kialakulhat a társadalmilag optimális szolgáltatási szint, míg a belépési díjak biztosítják, hogy a szolgáltatás ne legyen veszteséges. Ilyen módon a megnövekedett társadalmi többletből finanszírozható az adatszolgáltatás, mint ezt a következő példa is illusztrálja.

Egy állampolgár peres ügyéhez három hasonló bírósági eset már lezajlott a múltban. Amennyiben ezeket tanulmányozza, akkor egyre kedvezőbb képet kaphat esélyeiről, de az újabb eset megismerésének hozadéka csökkenő. Ha csak egy esetet tanulmányoz, többlethaszna 10 pénzegység (az egység legyen ismét ezer forint), ha még egyet, további többlethaszna 5, ha a harmadikat is, akkor újabb 1 többlethaszonhoz jut. Tegyük fel, hogy az adatszolgáltató számára egy esethez való hozzáférés biztosításának határköltsége 1, viszont elvárás, hogy ügyfélen legyen legalább 8 pénzegység nyeresége, amiből az adatbázis előállításának költségeit fedezik.

Ha csak egységarat alkalmazhatna, optimális számára az 5 ár választása, mivel ekkor az ügyfél 2 esetet kér le, ezért 10 pénzegységet fizet, profitja tehát $10 - 2 = 8$ egység. Ugyanakkor ha megengedjük, hogy a szolgáltató alkalmazzon egy belépési díjat, akkor az 1 egység ár választása mellett az ügyfél 3 esetet kér le, és

ezáltal $10 + 5 + 1 - 3 = 13$ többletet realizálhatna. A szolgáltató tehát választhat bármilyen 0 és 13 közötti belépési díjat, és az ügyfél számára ezt még mindig optimális választania, azaz a szolgáltató akár túl is teljesítheti bevételi célját.

A kétrészes tarifák alkalmazása igen elterjedt közszolgáltatások (telefon, gáz, stb.), illetve szellemi termékek árazásánál is (például copyright-szerződések). Adatszolgáltatásnál jellemző formája, hogy az adatkérő kifizeti az adatbázis használatának úgynevezett licenszdíját, majd meghatározott ideig ingyenesen kérhet le korlátlan számú adatot.

5.4. Többszintű árazási problémák

Az adatszolgáltatás árazásához kapcsolódó probléma, hogy a legtöbb esetben az adat iránt olyan felhasználók támasztanak keresletet, akik azt valamilyen termelési vagy közvetítési folyamat inputjaként használják fel, és előállítanak belőle egy újabb terméket. Ilyen gazdasági szereplők az információközvetítő cégek (például média), adatbáziskezelők illetve a különböző tanácsadó funkciókat ellátó vállalkozások (például ügyvédi irodák).

Az adatszolgáltató által meghatározott ár tehát valamilyen módon beépül az azt felhasználó piaci szereplők által meghatározott árba. Piacszerkezettől (például a piaci szereplők számától) függetlenül igaz az állítás, miszerint minél nagyobb az adatszolgáltató által meghatározott (input)ár, annál nagyobb lesz a végső termék ára, bár a fogyasztók felé áthárított költség nagysága már függ az adott piacszerkezettől. A magasabb ár csökkenti a végtermék piacán végbemenő tranzakciók nagyságát, így magát a társadalmi jólétet is. Amennyiben tehát valamilyen módon rákényszerítjük az adatszolgáltatót arra, hogy határköltség felett árazzon, akkor ez nemcsak a közte és az adatot „első körben” felhasználók közötti piacon vezet majd torzításhoz, hanem ez a torzítás továbbgyűrűzik azokra a piacokra, amelyeken ezt az adatot inputként használják.

A problémát jól érzékelteti annak anekdotikus formában megfogalmazott változata: a piaci hatalommal rendelkező cégnél csak egy rosszabb dolog van, a piaci hatalommal rendelkező cégek láncolata. Ennek az eredménynek a tükrében határozottan ellenjavallt az is, hogy a költségvetési szervek egymás között piaci alapon meghatározott egységáron cseréljenek egymás között információkat, valamint ez ellen szól az is, hogy a bürokrácia útvesztőiben ez a fajta holtteher-veszteség tovább duzzadhat.

Az ilyen többszintű piacokon különösen ajánlott tehát határkölség-alapú adatszolgáltatást alkalmazni, a fix költségeket pedig a kétrészes tarifáknál megismert licenszdíjakból fedezni. Az alacsonyabb határkölség alapú árazásnak az is eredménye, hogy az adatot inputként felhasználó vállalatok piacára könnyebb a belépés, ami erősebb versenyhez, ezáltal alacsonyabb árhoz és nagyobb jóléthez vezethet.

5.5. Társadalmilag optimális piacszerkezet az adatszolgáltatásban

Mint láttuk, az adatszolgáltató szervezet speciális költségstruktúrával rendelkezik: az adatbázis előállítás és üzemeltetése egy nagy egyszeri kiadást jelent, de ezek után minden újabb adatkérés teljesítése már csak egy állandó (ráadásul a fix költséghez képest relatíve kicsi) összegbe kerül. Az adatszolgáltatás átlagkölsége tehát annál alacsonyabb, minél nagyobb mennyiséget kínál a szervezet, vagyis a költségek annál nagyobb mértékben térülnek meg, minél nagyobb az általa szolgáltatott adattömeg.

Hatékony-e több, egyszerre adatgyűjtő és adatszolgáltató feladatot ellátó intézményt működtetni? Versenypolitikai érvelés alapján azt is mondhatnánk, hogy igen, hiszen a több szervezet között kialakuló versenytől jogosan várhatjuk el azt, hogy az alacsonyabb árhoz és ezáltal nagyobb jóléthez vezet. Költséghatékonysági szempontok ugyanakkor ellene szólnak több ilyen szervezet létrehozásának. Ennek oka egyszerűen az, hogy a határkölség konstans volta miatt mindegy, hogy hány szervezet látja el a szolgáltatást, hiszen az adatszolgáltatás folyamatát (vagyis egy

újabb adat hozzáférhetővé tételét) tekintve ugyanolyan hatékonyan tudnak működni, ugyanakkor több szervezet esetében többször ki kellene fizetni az adatgyűjtés (illetve az adatszolgáltató technológia kialakításának) nagy fix költséget, ami csökkenti a társadalmi jólétet. Az ilyen iparágakat a közgazdaságtan szemléletesen *természetes monopóliumnak* nevezi, mivel hatékonysági szempontból indokolt mindössze egyetlen szervezetnek helyet biztosítani a piacon. Adatgyűjtés területén tehát hatékony egy cég jelenléte, adatszolgáltatási és árazási politikáját viszont szabályozni kell, miként ez már az előző alfejezetekben is kifejtettük.

Jelenti-e ez az eredmény, hogy egyetlen szervezetnek kell elvégeznie mind az adatgyűjtő, adatszolgáltató és adatfeldolgozó feladatokat? Nem szükségszerűen. A nagy fix költségek az adatgyűjtésben jelentkeznek, ahol mint láttuk, indokolt a koncentráció. Nyilván ennek a szervezetnek is kell először az adatot szolgáltatnia, ugyanakkor mivel az információs jóságok reprodukálási költsége gyakorlatilag költségmentes, nem jelent hatékonyságvesztést, ha a felhasználók ezt továbbadhatják.¹⁴⁵ Az adatok feldolgozása esetében a fix költségek már kisebbek és a határköltség is valószínűleg növekszik a feldolgozott mennyiség függvényében. Az adatfeldolgozás területén tehát helye van több adatfeldolgozó szervezetnek is, és ezen a piacon inkább alkalmazhatóak a versenypolitika eszközei.

Az optimális piacszerkezet tehát az adatgyűjtés és adatszolgáltatás „felső” szintjén monopólium létét tartja szükségesnek, amely nyilvánosságra hozza az adatait, az adatfeldolgozás „alsó” szintjén pedig megengedi a verseny meglétét. Az előző fejezetben mondtak szerint ugyanakkor hatékony, ha az elsődleges adatszolgáltató határköltség-alapú árazást alkalmaz (amely ár közel nullának tekinthető), ami elérhető akár direkt árszabályozással, akár kétrészes tarifák engedélyezésével. Az ilyen módon kialakult versenyhelyzettől azt is várjuk, hogy a feldolgozott adatok végső felhasználóinak igényeit jobban figyelembe veszik, ami tovább növelheti a jólétet. Ezen felül az adatokhoz határköltségen való hozzá hozzáférés, azaz gyakorlatilag az adatok nyilvánossága azt is lehetővé tenné, hogy a verseny minőségi kontrollt biztosítson.

¹⁴⁵ Az Alkotmány ezt lehetővé is teszi: a közérdekű adat nemcsak abban az értelemben nyilvános, hogy bárki hozzáférhet, hanem azt bárki „újra” nyilvánosságra is hozhatja, azaz terjesztheti.

5.6. Az adatszolgáltatás okozta externáliák és ezek figyelembevétele az árazásban

Eddigi megközelítésünk során az adatszolgáltatás hasznait a nettó fogyasztói többlettel mértük, vagyis összeadtuk, hogy a fogyasztók számára *egyénilag* mennyit ér az adott adat elérése (mekkora összeget adnának érte) és kivontuk belőle azt az összeget, amennyit ténylegesen fizetnek érte. Ezen számítási mód alapján bemutattuk, hogy társadalmilag optimális az adatokat azok reprodukálásának határköltségén elérhetővé tenni, illetve amennyiben erre nem állnak rendelkezésre források, milyen diszkriminatív ártorzításokat érdemes alkalmazni.

Ez a megközelítés arra a feltevésre épül, hogy egy gazdasági szereplő az őt érintő egyéni hasznok és költségek alapján hozza meg döntéseit. Léteznek ugyanakkor olyan termékek, amelyek fogyasztása révén egy adott szereplő más gazdasági szereplők jólétét is befolyásolja, úgynevezett externális hatással van. Az adatokhoz való hozzáférés tipikusan pozitív externáliának tekinthető, mivel az informált szereplők megszerzett tudásuk bizonyos részét költségmentesen átadják a gazdaság más szereplőinek, növelve ezzel az ő informáltságukat, és ezúttal jólétüket. Egy lakóközösség tagjai például széles körben cserélnek a helyi önkormányzat szolgáltatásaira (például ügyfélfogadási rendjére) vonatkozó információkat. Még szembetűnőbb pozitív externália képződik tudományos tevékenység révén, amit a kutatók a saját érdekeiket követve folytatnak, eredményeik mégis széles körben hasznosulnak.

Az externália kifejezést az indokolja, hogy ezeket a hatásokat az adott szereplő nem internalizálja, vagyis nem veszi figyelembe döntései meghozatalakor.¹⁴⁶ Ez a tulajdonság számos társadalmilag hatékony döntés meghozatalát gátolhatja: amennyiben például egy adatkérés döntéshozónál jelentkező költsége (az adat ára) nagyobb, mint a döntéshozó informáltságából fakadó egyéni haszna, ugyanakkor

¹⁴⁶ Ilyen értelemben az a hatás, ami akkor jelentkezik, ha egy vállalat új munkahelyeket teremt, hogy az általa hozzáférhető adatokat feldolgozza, nem externális hatás, mivel a dolgozói által létrehozott érték a vállalat profitját gyarapítja.

kisebb, mint a döntés informáltságából fakadó hasznok összessége, akkor a döntéshozó eláll az adatkéréstől, mivel az számára nem megtérülő befektetés. Sok esetben ez a jelenség szolgálhat magyarázatul az adatok nyilvánosságra hozatalával szembeni ellenállásra olyan adatfeldolgozó intézmények esetében, amelyeknek az adatszolgáltatás nem elsődleges feladata, mivel nem ismerik fel az adatfelhasználásból származó externális hatásokat.

Mit mondhatunk pozitív externáliák esetén a társadalmilag optimális árazásról? Mint az előző bekezdésben szereplő példa mutatja, pozitív externáliával bíró termék esetében a határkötség-alapú árazás sem biztosítja, hogy minden társadalmilag optimális tranzakció megvalósuljon. Ahhoz, hogy a kereslet az optimális szintre növekedhessen, az árat hatékony a határkötség alatti szinten meghatározni. Bár a nyilvánosság által biztosított nulla ár nem csökkenthető tovább, a hozzáférés költségeit egyéb módokon is lehet csökkenteni, amelyekre a következő fejezetben láthatunk példákat.

A fejezet középső részében bemutattuk, hogy az adatszolgáltatás speciális költségstruktúrája és a finanszírozási források szűkössége indokoltta teheti a határkötség-alapú árazástól való eltérést. Mint láttuk, amennyiben az adatigénylők köre jól elkülöníthető különböző árugalmasságú fogyasztói csoportokra, érdemes velük szemben diszkriminatív árazást alkalmazni. De milyen szabályozott árakat válasszunk, amennyiben az adatfogyasztók különböző mértékben fejtenek ki pozitív externáliákat a társadalom többi tagjának irányába?

A jólét-maximalizáló árdiszkrimináció tárgyalásánál megjegyeztük, hogy az annál érvényesített árdiszkriminációs irány megegyezik azzal, amelyet egy profit-maximalizáló szolgáltató választana, vagyis a kevésbé árérzékeny felhasználói csoport irányába szabnak nagyobb árat. Externáliák esetében ugyanakkor ez az állítás immár nem áll fent, ugyanis egy profit-maximalizáló szolgáltató nem tudja „beárzni” az ügyfélkörénél keletkező externális hatásokat, mivel azokat maguk a fogyasztók sem internalizálják. Ebben az esetben tehát a regulációnak nemcsak az esetleges árakat kell arányosan csökkenteni arra a szintre, hogy a bevételek éppen

fedezzék a szolgáltató költségeit, hanem azok struktúráját is meg kell változtatni. Azokat a szereplőket, akik relatíve nagyobb pozitív externális hatást fejtenek ki („közhazsnúbb” tevékenységet végeznek), további támogatásban kell részesíteni, vagyis kisebb árat meghatározni az irányukban, ami természetesen azt is jelenti, hogy a relatíve kisebb vagy semmilyen externális hatást okozó fogyasztóknak megszabott árat kell növelni. Ez a fajta racionalitás is meghúzódik az egyes adatszolgáltató-szervezetek által alkalmazott árazási alapelvek mögött, például külföldi szereplők számára általában drágább bizonyos adatokhoz való hozzáférés, illetve kutatási célokra kért adatok könnyebben hozzáférhetőek.¹⁴⁷

Hangsúlyozni szeretnénk, hogy bár tisztán látható, hogy az adatszolgáltatás pozitív externális hatásokat okoz, ezek számszerűsítése igen nehéz feladat, mivel az externália nagyságára kibocsátójának viselkedéséből erre definíciószerűen nem következtethetünk, ugyanis az egyéni döntést ez nem befolyásolja. Alkalmazhatóságának nehézsége miatt erre a megközelítésre inkább csak érintőlegesen támaszkodunk, amikor két felhasználói módot összehasonlítva erősen valószínűsíthető, hogy melyik tekinthető „közhazsnúbbnak”.

¹⁴⁷ Az utóbbi esetben ez sokszor nem jelenik meg tényleges különböző árak formájában, de kutatási célok esetében az adatszolgáltató hajlamosabb felismerni a döntése által okozott pozitív externáliát, és ezáltal jobban együttműködik az adatkérővel, amivel csökkenti az adatkérés nem pénzben jelentkező költségeit.

6. A klasszikus közérdekű adatokhoz való hozzáférés hasznai és költségei

Ebben a fejezetben a klasszikus közérdekű adatok felhasználásából eredő hasznokat és költségeket tekintjük át. A klasszikus adatokat a tanulmány elején úgy definiáltuk, amelyeket fogyasztók közvetlenül felhasználnak, hogy jobb döntéseket hozhassanak. A fejezet első részében részletesebb betekintést nyújtunk egy közgazdasági modellbe, ami megmagyarázza az információ iránti kereslet kialakulását, és rávilágít arra is, az adathozzáférés költségei hogyan befolyásolják a fogyasztók által igényelt adat mennyiségét.

Minden jel arra mutat, hogy az ezen adatok hozzáféréséből származó hasznokat a politikai döntéshozók felismerik, és kellően nagynak gondolják. Ennek eredményeként a klasszikus közérdekű közül a legtöbb hozzáférhető, ezen belül pedig igen sok ingyen, vagy mérsékelt pénzkiadás árán. Ugyanakkor az esettanulmányokon keresztül bemutatjuk, hogy ha csak egy bizonyos módon teszik ingyenesen hozzáférhetővé ezeket az adatokat, attól még nem minden felhasználó számára lesz a hozzáférés ténylegesen költségmentes. Sok esetben azt is látjuk, hogy bár a közel költségmentes hozzáférési módszer létezik, a létezéséről beszámoló amúgy szintén közérdekű adat nem érhető el költségmentesen, ami szintén akadályt állít a hozzáférés elé.

Pár esettanulmány esetében arra is rámutatunk, hogy amennyiben a közérdekű adatokat egy piaci vállalkozás számára is rendelkezésre bocsátják, az sok esetben hatékonyabb információszolgáltatói feladatokat lát el, mint az állami szervezetek. Bár ennek bizonyos szolgáltatások esetében nagyobb az ára, több példát látunk arra, hogy az alapszolgáltatások esetében megvalósul az ingyenes hozzáférhetőség. Az üzleti vállalkozások ugyanakkor sokkal nagyobb hangsúlyt fektetnek az információforrás létére vonatkozó információk terjesztésére, hiszen közvetlenül érdekeltek az információ fogyasztásának terjedésében, így hatékonyabb információszolgáltatóként működhetnek.

6.1. Az információkeresés alapmodellje

Vizsgáljuk meg először kicsit alaposabban a klasszikus közérdekű adatok iránti kereslet kialakulását! Arra a kérdésre keressük a választ, hogy miért is van szükségük az embereknek információra és mennyi információt áll érdekükben megszerezniük. Ezt a gondolatmentet azért hangsúlyozzuk részletesebben, mert – ha eltekintünk az esetleges jogi kötelezettségektől – ez az, ami alapján az adatgazda, illetve az állam meghozza azt a döntést, hogy mely adatok nyilvánosságát teremtsen meg, és milyen áron szolgáltatassa az adott információt.

Egy felhasználó számára az adatokban rejlő információt azért tekintjük hasznosnak, mert a több megismert lehetőség közötti választás lehetősége növeli a várhatóan elérhető hasznokat, illetve csökkenti a várható költségeket. Az információ tehát csökkenti a döntéseink következményei felőli bizonytalanságot, és ezért cserébe hajlandóak vagyunk bizonyos árat fizetni. Ennek az állításnak a szemléltetésére tekintsünk egy egyszerű példát, amit a közgazdaságtanban az információkeresés klasszikus modelljének tekintenek.¹⁴⁸

Tegyük fel, hogy egy nagyobb értékű terméket akarunk megvásárolni, amelynek piaci árát 200 és 220 ezer forint közé becsüljük. Az egyszerűség kedvéért tegyük fel, hogy minden üzletben 50 százalékos eséllyel 200, 50 százalékos eséllyel 220 ezer forint a termék ára, de előzetesen nem tudjuk, melyik üzletben mekkora árra számíthatunk. Feladatunk tehát az, hogy megtaláljuk a terméket legalacsonyabb áron kínáló üzletet. Ha a keresés költségmentes lenne, akkor a probléma egyszerű: be kell menni minden esélyes üzletbe. Ha azonban a keresésnek költsége van, a fogyasztónak előzetesen el kell döntenie, hogy hány üzlettől kér árajánlatot, amiből majd kiválasztja a legkisebbet. Ha minden adatlekérés költsége 2000 forint, akkor a kérdés az, hogy hány ajánlatot kell kérnie, mekkora adatmennyiséget optimális beszereznie? Ezt a kérdést az információ beszerzése előtt várható költségek és hasznok mérlegelésével lehet eldönteni.

¹⁴⁸ A gondolat első megjelenésére lásd Stigler (1961).

Az egyszerűség kedvéért tegyük most fel, hogy csak két ár létezik: a legmagasabb (220) és a legalacsonyabb (200). Amennyiben az érdeklődő csak egy ajánlatot kér, az ár várható értéke $0,5 * 200 + 0,5 * 220 = 210$ ezer forint, hiszen a bolt egyenlő eséllyel kínálja a terméket a kisebb vagy a nagyobb áron. Amennyiben kettőt, akkor 25 százalék eséllyel mindkét ajánlat 220 ezer, vagyis a fogyasztó 220 ezret fizet a termékért, 75 százalékkal pedig legalább az egyik ajánlat 200 ezer, és a fogyasztó nyilván ezt az alacsonyabb árat választja majd.¹⁴⁹ Két adatkérés esetén az ár várható értéke tehát $0,75 * 200 + 0,25 * 220 = 205$ ezer forint, vagyis a második adatkérés addicionális, úgynevezett határhaszna az elsőhöz képest $210 - 205 = 5$ ezer forint. Hasonlóképpen kiszámolható, hogy ha három ajánlatot kér, az ár várható értéke $0,875 * 200 + 0,125 * 220 = 202,5$, négy ajánlat esetén 201,25 és így tovább. Mint látjuk, az újabb adatkérések határhaszna pozitív, de egyre kisebb: a harmadiké 2,5, a negyediké pedig 1,25 ezer forint. Meddig érdemes a fogyasztónak újabb ajánlatokat kérni? A harmadik adatkérés még kifizetődő, hiszen az adat megismerésétől várható haszon magasabb, mint a megszerzés határköltsége, ugyanakkor a negyedik már nem.

A fenti példában megmutatott tendencia általában is igaz: minden újabb adat megismerése pozitív határhaszonnal bír, de ez a határhaszon egyre inkább csökken. Egy racionális adatigénylőnek tehát csak addig a lekért adatmennyiségig érdemes elmennie, amíg az adatkérés határköltsége meg nem haladja az utolsó adat lekéréséből származó többlethasznót. Amennyiben a hozzáférés határköltsége nullánál nagyobb, egyénileg nem kifizetődő teljesen informáltnak lenni. Mindig marad tehát esély arra, hogy a keresés akár igen nagy anyagi befektetés után sem jár a kívánt eredménnyel, és a fogyasztó nem találja meg a legolcsóbb megoldást.

Ez a modell arra is jó választ ad, hogy bizonyos emberek miért rendelkeznek jobb, mások kevésbé jó (vagy egyszerűen: kevesebb) információval. Akinek kisebb

¹⁴⁹ Az olvasó számára talán furcsának tűnhet, hogy a fogyasztó akkor is tovább keres, ha már az első ajánlat a legalacsonyabb árat eredményezi. Ezzel a feltevessel csak azért élünk, hogy gondolatmenetünk matematikailag egyszerűen szemléltethető legyen, ha minden keresés után megengednénk, hogy a fogyasztó megálljon, az nem változtatná meg kvalitatív következtetéseinket. A feltételezés talán könnyebben elfogadható, ha azt tesszük fel, hogy az ajánlatokat valamilyen oknál fogva egyszerre érdemes, vagy lehet csak kérni.

haszna származik az informáltságból, vagy nagyobb számára a hozzáférés határkölsége, kevesebb információt fogyaszt. Az utóbbi ok igen gyakori lehet a hátrányos helyzetű társadalmi rétegek körében, akik számára a keresés valamilyen oknál fogva túl költséges. Az információszerzés költségének csökkentése az ő számukra járhat tehát a legnagyobb haszonnal.

Számos más mindennapi példát adhatunk, amelyek a szemléltető árkeresési példánkhoz hasonló jellemzőkkel bírnak:

- Törvényhozási gyakorlat: egy pereskedő fél minél több saját ügyéhez hasonló ügybe nyer betekintést, annál kisebb a bizonytalansága a saját ügyének kimenetele felől.
- Törvényességi aggályok: egy fogyasztó minél több forrásból meggyőződik arról, hogy az általa megvásárolni kívánt árú legális eredetű, annál nagyobb értéket tulajdonít neki, vagy ami ezzel egyenértékű, annál kisebb költségekkel számol birtoklása esetén.
- Életminőségi kérdések: egy gyermeket váró nő minél több kórház körülményeit, illetve a szüléssel kapcsolatos egyéb költségeket (például: hálapénz) ismer meg, annál jobban össze tudja vetni a kórház- és orvosválasztással várható hasznokat és költségeket. Hasonló a helyzet a gyermekes szülők iskolaválasztási problémája esetében.
- Meteorológiai adatok: egy gazdálkodó minél több információt szerez az elkövetkező hónapban várható időjárásról, annál hatékonyabban szervezheti meg mezőgazdasági tevékenységeit, amely költségmegtakarításokkal és / vagy bevétel-növekedéssel jár.

Tegyük fel egyelőre, hogy az adatok elérése senki számára sem prohibítív, azaz nincsenek olyan fix költségek, amelyek az adatelérést magát akadályoznák, így mindenki képes legalább a legkisebb mértékben elérni az adatokat! Hogyan növelhető ilyenkor az érintett szereplők által választott informáltsági szint? Az egyszerű válasz az, hogy csökkenteni kell az adatokhoz való hozzáférés határkölségét. Ezek után azonban még mindig ott marad a kérdés, megéri-e ez az intézkedés, és kinek kell ezt megtennie? Megválaszolásához tudnunk kell, hogy egy

bizonyos adat biztosítása mekkora hasznokat eredményez, amihez összegeznünk kell az érintett fogyasztóknál az adatfogyasztásból származó többlethasznokat és megtakarításokat.

Az előző példát folytatva tegyük fel, hogy 1000 fogyasztó szembesül az ott leírt árkeresési problémával, és létezik egy technológia, amely által az adatkérés határkölsége 1000 forintra csökkenthető. Amennyiben a technológia egyszeri 1 millió forintba kerül, társadalmilag hatékonynak tekinthető-e a bevezetése?

Az eredeti 2000 forintos adatlekérési díj mellett minden fogyasztó három adatot kért le, 1000 forintos díj mellett pedig négyet. Az új helyzetben tehát minden fogyasztó hozzájut a negyedik keresés 1,25 ezer forint értékű határhasznához, ami összesítve 1,25 millió forintot tesz ki. Mivel az összhaszon meghaladja az esetenként olcsóbb adathozzáférés kialakításának költségeit, ezért ennek a technológiának a bevezetése társadalmilag kedvező.

Ez a példa rámutat arra az általános problémára is, hogy az olcsóbb adathozzáférés kialakításának költsége általában túl magas ahhoz, hogy az érintett fogyasztóknak egyénileg megérné befektetni ebbe a technológiába. Ezt a szerepet tehát valakinek át kell vállalnia a fogyasztóktól.

Mindeddig elsősorban az adathozzáférés határkölségére koncentráltunk, azonban az adathozzáférés fix költsége sem kerülhető meg. A fix költség számunkra két okból érdekes. Az egyik az, hogy mértéke a társadalom egésze vagy egyes részei számára elrettentő lehet – olyan nagy, hogy az önmagában elérhetetlenné teszi az adatok elérését (ennek speciális esete az, ha egy adat nem is nyilvános). A másik kérdés akkor merül fel, ha – ahogyan azt a közgazdasági elméleti háttérrel tárgyaló fejezetben láttuk – az adathozzáférés feltétele (a példánkban szereplő helyzethez hasonlóan) egy technológia, és az azt finanszírozó intézménye kétrészes árazás bevezetésén gondolkodik. Ezekben az esetekben nem csak a keresés határkölségét, hanem a fix költségek csökkentését lehetővé tevő intézkedéseket is fontolóra kell venni.

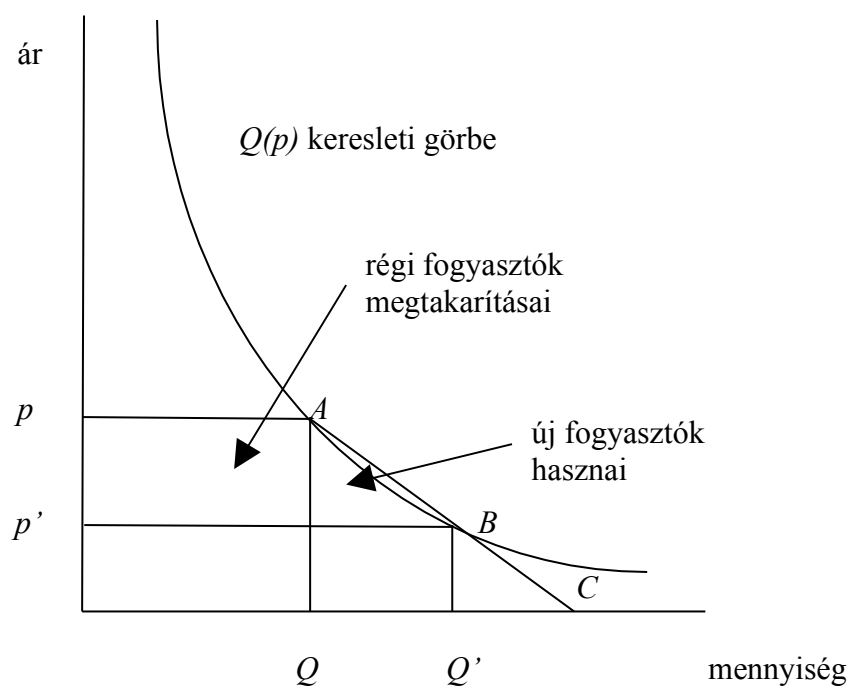
Számos piaci példát ismerünk, amikor piaci szereplőknek érdekük fűződik ahhoz, hogy az információt olcsóbban jutassák el a fogyasztókhoz, és ezért költségeket is hajlandóak vállalni. Ez a kezdeményezés általában a következő két forrásból eredhet:

- Olyan eladóktól, akik a fogyasztó számára kedvező tulajdonsággal bírnak, mint például alacsony ár (diszkontüzletek), vagy jó minőség (márkás termékek). Ezek a szereplők tisztában vannak azzal, hogy ha a fogyasztóhoz eljut a rájuk vonatkozó információ, akkor ezáltal megnő a forgalmuk, és az így elért bevétel-növekedés kompenzálja az információszerzés megkönnyítésének (például reklámozás) költségét.
- Információközvetítő szereplőktől, akik az eljuttatott információért cserébe igényt formálnak a fogyasztó abból származó hasznának egy részére. Ilyenek például a különböző jogi vagy piaci tanácsadó cégek. Szemben az előző csoporttal, ezek a szereplők jelentős részben közérdekű adatokkal vagy az azokból levezethető információkkal kereskednek.

A példáinkban felsorolt adatokkal azonban eredendően nem vállalkozások, hanem az állam vagy annak intézményei, illetve a közfeladatot ellátó más szervek rendelkeznek, ezért is minősülnek közérdekű adatnak. Ez történhet azért, mert azok speciális jogi szerepet töltenek be, mint a telekkönyvi adatok, vagy valamilyen formában érzékenyek, mint az egészségügyi intézmények helyzete. De igen sokszor (ugyanakkor nem minden esetben) az ok pusztán annyi, hogy a létrehozásuk költségei annyira nagyok, hogy azt a hazai viszonyok között csak az állam képes vállalni – ilyenek például a meteorológiai adatok. A fejezetünk második részében ezeken a példákon keresztül tekintjük át, az állam hogyan segíti a hozzáférési költségek csökkentését, és ezzel az állampolgári döntések hatékonyságának növelését.

6.1.1. Hasznok számszerűsítése

A részletesebb tárgyalás előtt vázoljuk még fel, hogy milyen adatokra lenne szükségünk ahhoz, hogy számszerűsítsük egy adott költségcsökkentő intézkedésből származó hasznokat, amelyek összevethetőek lennének ezen költségcsökkentő beruházás költségeivel. Tegyük fel, hogy egy adott adat megszerzésének költsége p (ami lényegében a hozzáférés nem csak pénzben mért ára), és az adat szolgáltatójától rendelkezésünkre áll az a forgalmi adat, hogy ezen ár mellett Q számú adatigénylés történt. Ezen adat alapján megkapjuk az adat iránti keresleti görbe egyik pontját, jelöljük ezt az alábbi ábrán A-val!



5. ábra

Ha az ár p' -re csökken, kétfajta nyereség képződik. Egyrészt akik a magasabb áron is igényelték volna az adatot, azok továbbra is megteszik ezt az alacsonyabb ár mellett, vagyis minden eddig adatigényléshez a $(p - p')$ árkülönbségnyi költségmegtakarítás tartozik, ami összesen $(p - p') * Q$ nyereség a régi fogyasztóknak. Másrészt az alacsonyabb ár mellett újabb fogyasztók számára is megéri igényelni az adatot. Ennek a hatásnak a számszerűsítéséhez ugyanakkor ismernünk kell azt, hogy a p' ár mellett mekkora a Q' kereslet nagysága. Tegyük fel, hogy ez az árváltozás valóban megtörtént, és az adatszolgáltató rendelkezésünkre bocsátja ezt a forgalmi adatot is, jelölje ezt az ábrán B . Mint ezt az előző közgazdasági fejezetben megmutattuk, az új fogyasztók nettó haszna ekkor a keresleti görbe és az új p' ár közötti terület nagysága. Amennyiben a keresleti görbe az adott szakaszon jól közelíthető egy egyenessel, akkor ezen további hasznok értéke megközelíthető azon háromszög területével, amelynek két befogója az árváltozás és a keresletváltozás. A pótlólagos haszon ennek megfelelően

$$\frac{(p - p') * (Q' - Q)}{2}.$$

Ezen két forgalmi adat ismeretében adhatnánk egy durva becslést arra nézve is, milyen nyereségek képződnének az ár nullára csökkentésével. Amennyiben csak az A és a B pont áll rendelkezésünkre a keresleti összefüggés megbecsléséhez, akkor azokat összekötve és meghosszabbítva a C ponttal becsülnénk az ingyenesség melletti hozzáférés nagyságát, és a fentiekhez hasonló módon számolhatnánk ki a megfelelő területek nagyságát.

Természetesen minél több forgalmi adatunk van az adatszolgáltatásra nézve, annál pontosabb becslést adhatunk. Itt kifejezetten azt akartuk megmutatni, hogy már igen korlátozott mennyiségű adatból is adhatók becslések a keletkező hasznok számszerűsítésére. Sajnos a legtöbb esetben még egy vagy két forgalmi adat beszerzése is igen problémásnak bizonyult, így ezeknek az egyszerű becsléseknek az elvégzése is problémákat okozott. Ezen becslési eljárás egyszerűsége rámutat

arra is, hogy az adatszolgáltató intézményeknek nagyobb figyelmet kell szentelniük forgalmuk dokumentálására. Teljesítményük hasznai így indirekt módon könnyedén számszerűsíthetők lennének, ehhez azonban elengedhetetlenek a megfelelő (szintén közérdekűnek minősülő) adatok.

6.2. Szelektív áttekintés a közérdekű adatokhoz való hozzáférés lehetőségeiről Magyarországon

A közérdekű adatok nyilvánosságának kérdése a közgazdasági részben kifejtett elemzési, és a fent leírt mérési keretben tulajdonképpen triviális. Az adatok az állam kezében vannak, nyilvánosságra hozataluk költsége relatíve csekély, használatuk hasznot hajt. Az alábbiakban nem is azt vizsgáljuk, hogy optimális-e nyilvánosságra hozni őket, mivel erre a válasz nyilvánvalóan igen, inkább arra keresünk választ, hogy a már nyilvános adatok esetében milyen a nyilvánosság mértéke, milyen költséggel férnek a felhasználók az adatokhoz, és az adatgazdák milyen árazási stratégiákat alkalmaznak. Ezen ismeretek szükségesek ugyanis ahhoz, hogy meg tudjuk ítélni a nyilvánosság jelen mértékének hasznait, és véleményt tudjunk alkotni arról, hogy milyen hasznok várhatók egy, a jelenleginél nagyobb nyilvánosságot teremtő helyzetben.

Az adathozzáférés fogyasztói oldalának elméleti vizsgálatakor hangsúlyoztuk: az adathozzáférés költségein nem csak annak pénzben megjelenő költséget értjük, hanem minden, a megszerzéshez szükséges erőforrás költségét. Egy ilyen jellemző dimenzió az adatokhoz való hozzájutás időigényének, vagy bizonyos hozzáférési módokhoz kapcsolódó egyszer fizetendő, ugyanakkor általában nagy összegű beruházások (például Internet- vagy mobiltelefon-hozzáférés, kapcsolatok kiépítése) figyelembevétele. Tekintve, hogy a hasznok sok esetben kézenfekvőek, vizsgálatunk most következő, esettanulmányokra épülő szakaszában ezt a kérdést domborítjuk ki. Mivel egyes adatkéréseknél különböző költségek és ebből fakadóan különböző problémák jelentkeznek, ezért ezek alapján tovább tipizáltuk a klasszikus közérdekű adatokat, és minden típus gyakorlatát egy-egy magyarországi példán keresztül vizsgálunk meg. Bár törekszünk általános érvényű állítások megfogalmazására,

inkább elemzési módokat mutatunk be, amiket a típuseseteknél lehet alkalmazni, és nem állítjuk, hogy minden azonos típusba eső adathozzáférésnél azonos eredményt kapunk.

6.2.1. Keresési lehetőségek megismerése

Az állampolgári adatkéréseknek sokszor szab gátat az, hogy az adatigénylők nem tudják, kitől és milyen módon szerezhetik be a kívánt adatot. Első vizsgált területünk tehát az adatforrásokra vonatkozó adatokhoz való hozzáférés lehetősége, például hogy hogyan tájékozódhatunk egy adott önkormányzat ügyfélfogadási rendjéről.

Az ilyen adatokhoz való hozzáférés haszna jellemzően abból ered, hogy a fogyasztó időt takarít meg, bár bizonyos esetekben előfordulhatnak konkrét pénzbeli megtakarítások is (például megtudjuk, hogy az adathoz való hozzáférésnek van egy ingyenes módja is a közismert költségtérítéses változattal szemben, erre látunk példát a későbbiekben a gépkocsi-nyilvántartásnál). Mik az ilyen típusú adatokhoz való hozzáférés költségei? A fogyasztónak időt kell szánnia az adatforrások keresésére, valamint hozzá kell férnie a keresésre lehetőséget biztosító infrastruktúrához. A keresés költségeit tehát úgy csökkenthetjük, hogy lerövidítjük a kereséshez szükséges időt, illetve támogatjuk a kereső technológiák széles körben való elterjedtségét.

Igen jellemző módszer a nyilvánosság biztosítására az adatok Interneten való közzététele. Ez a tény azonban még nem biztosítja, hogy a fogyasztó automatikusan rá is talál a kért adatra. A keresés időigényességét például a következő módokon próbálhatjuk közelíteni:

- Kulcsszóra keresés esetén hányadik helyen találjuk az adatszolgáltató forrást? A keresési helyezés előállításának de facto standardja ma a Google internetes kereső.

- Létezik-e valamilyen központi portál, amely eligazítást nyújt az adatszolgáltatókról? Ilyen központi portál például a magyarorszag.hu, amelyről lentebb röviden írunk.
- Mennyire áttekinthető az adott adatszolgáltató honlap? Ezt a kérdést vizsgálja például megyei jogú városok önkormányzati honlapjának összehasonlító elemzésével az Eötvös Károly Intézet Elektronikus adatnyilvánosság c. tanulmánya.
- Létezik-e valamilyen hozzáférést megkönnyítő rendszer? Erre látunk példát az Ügyfélkapu rendszer áttekintésével.¹⁵⁰

Az ilyenfajta tájékoztató adatok Internetes közzététele természetesen nem jelent egyetemes megoldást, hiszen az Internethez hozzáférő réteg az állampolgárok nem túl nagy arányát teszi ki, és ezen belül is túlreprezentáltak a magasabb jövedelmű társadalmi csoportok. Magyarországon az Internet-penetráció 20-25%-os, a munkahelyeken vagy oktatási intézményekben Internethez hozzáférőkkel együtt is csak 35%.¹⁵¹ Ennek részben az az oka, hogy az egyéni Internethez való hozzáférésnek nem elhanyagolható fix költségei vannak: a saját számítógép bolti ára minimum 70 ezer forint, az Internet-csatlakozás előfizetés sebességtől függően 0 – 15000 forint/hó. Habár egyszeri beruházás esetén az újabb adatkereséseknek gyakorlatilag semmilyen addicionális költsége nincsen, igen valószínű, hogy pusztán adatforrások felkutatására kevesen vásároltak számítógépet.

Helyi önkormányzatok és más szervek esetében is számos példát láthatunk, hogyan biztosítanak helyettesítő lehetőségeket ezen információk eljuttatására az állampolgárokhoz. Ilyen például az internetes hozzáférési pontok biztosítása a polgármesteri hivatalokon belül vagy e-Magyarország pontokon, telefonos ügyfélszolgálatok működtetése, az ingyenes tájékoztató kiadványok. Ezeknek az adatigénylésekhez kapcsolódó költségek szerkezete különbözik az előzőtől, mivel a

¹⁵⁰ Az adatok fellelhetőségét szolgáló újítás az elektronikus információszabadság-törvény alapján létrehozandó közadat-kereső rendszer is, amelyről A közszférában kezelt adatok típusai a magyar jogrendszerben című tanulmány szól röviden.

¹⁵¹ Az NRC online piackutató cég beszámolóját a magyarországi internetes helyzetről tekinti át a Népszava 2005-11-03-n megjelent „Kattanj a világra” c. cikke: <http://www.nepszava.hu/default.asp?cCenter=article.asp&nID=762344>

használatukhoz szükséges technológia kisebb fix költséget jelent, ugyanakkor igénybevételükkor jelentősebb változó (jellemzően időbeli) költségek jelentkeznek.

Az Internet széles körben való elterjesztését Magyarországon is fontos állami feladatnak tartják, és számos ezt támogató programot dolgoztak ki, a legismertebb ezek közül a Sulinet. Bár a támogatott szolgáltatások köre is sok ponton vitatható,¹⁵² az alapvető probléma az, hogy az Internet-használat terjedése igen lassú. Az Internet használatának képessége – elvonatkoztatva a szükséges technikai eszközök birtoklásától – korántsem éri el azt a szintet, amilyen természetességgel a legtöbb ember például a telefont kezeli.

6.2.1.1.A magyarorszag.hu és az Ügyfélkapu-rendszer példája

Az elektronikus kormányzati programon belül 2001-ben hozták létre egy 400 millió forintos beruházás keretében a magyarorszag.hu kormányzati portált. Ez a portál áttekintést nyújt a magyarországi közigazgatás főbb hivatalairól, valamint lehetőséget biztosít egyes közigazgatási ügyek elintézésére. Az utóbbi lehetőségek azonban csak úgy vehetőek igénybe, ha az ügyfél regisztrálja magát az Ügyfélkapu-rendszerben, amely a magyar kormányzat elektronikus ügyfélbeléptető és azonosító rendszere.

Amikor az érdeklődő az Ügyfélkapun először „átlép” – hacsak nem rendelkezünk a meghatározott minősítésű elektronikus aláírással – át kell esnie egy azonosítási eljárásnak: az Internet-használat képességén túl az erre fordított idő az Ügyfélkapu használatának költsége (az első regisztráció ingyenes). Az azonosítás bármely okmányirodában lefolytatható, és tényleges időtartama nem haladja meg az 5 percet (előzetes időpontfoglalás esetén, amely telefonon vagy Interneten is megtehető, a rendszer két napon belül ad időpontot). Mindezt megelőzheti az Interneten keresztül történő ideiglenes regisztráció, amely valódi

¹⁵² Ilyen különösen ellentmondásos rész a SuliNet Expressz program, amely kiegészítő elektronikus eszközök (például digitális kamera) vásárlását támogatta, amelynek kevés köze volt az alapszolgáltatás felhasználói körének kiterjesztéséhez. Az Internet-előfizetés költsége ezzel szemben nem visszaigényelhető.

költségmegtakarítást nem jelent, mert a szolgáltatásoknak csak igen szűk köre vehető igénybe.

Végleges regisztráció esetén az ügyintézés gyorsításával elérhető hasznok jelentősek. Elérhetővé válnak az Adó- és Pénzügyi Ellenőrzési Hivatal szolgáltatásai (például elektronikus adóbevallás, adófolyószámla-kivonat igénylése, kontroll adatszolgáltatások), okmányirodai szolgáltatások (anyakönyvi kivonat és lakcímgigazolvány igénylése, egyéni vállalkozói igazolvánnyal és járműigazgatással kapcsolatos ügyintézés, időpontfoglalás okmányirodába) és a Kormányzati Portál szolgáltatásai (gépjárműkereső, cégkereső, ingatlankereső).

6.2.2.Közérdekű tények megismerése

Amikor közérdekű adatok felhasználására gondolunk, legtöbbször az a példa juthat eszünkbe, amikor valaki egy állami szervezettől egy közérdekű ténnyt igényel, és ennek megismerése után hozza meg döntését. Tipikus példa lehet az az eset, amikor az állampolgár, mielőtt dönt egy ingatlan megvásárlásáról, kikéri annak tulajdoni lapját, hogy megbizonyosodjon a tulajdonos személyéről. Ezen közérdekű tények nyers formájukban hasznosak az adatkérő számára, mivel egy egyszerű kérdésre adnak könnyen értelmezhető választ.

6.2.2.1.A cégnyilvántartás példája

A magyarországi cégek nyilvántartását a cégjegyzékben összesítik. A gazdasági társaságokról szóló törvény rendelkezése szerint a cégjegyzék adatai nyilvánosak, ezeket a Céglélap c. lapban teszik, amely az Igazságügyi Minisztérium hivatalos lapja. A Céglélap 2004-től csak elektronikus kiadásban (CD-n) jelenik meg, előfizetési díja egy évre 112 ezer forint. Alternatív módszerként egyedi adatkérésekre ingyenes hozzáférést biztosít a magyarorszag.hu weboldal Ügyfélkapus regisztráció után, 10 lekérdezést biztosítva havonta.

Ezt az információ ugyanakkor nem csak állami szervezettől szerezhető be. A piaci alapon működő KJK-KERSZÖV Jogi és Üzleti Kiadó Kft. által működtetett www.complex.hu honlapon korlátlan mennyiségű egyedi adatkérést végezhetünk. Ez utóbbi szolgáltatás ugyanakkor csak havonkénti adatbázis-frissítést ígér.

A cégekre vonatkozó adatok összességét tartalmazó adatbázis csak ettől az utóbbi információközvetítő-szervezettől szerezhető meg, Complex Céghírek néven (CD-n illetve DVD-n, évi 70 ezer forint előfizetési díj ellenében). Ez a havonta frissülő, közel 20 évre visszatekintő adatbázis a Cégek Közlöny adatain alapul, számos keresési lehetőséggel, cégtörténettel, cégek közötti tulajdonosi kapcsolatok feltárásával, stb. Előfizetői részére további extra szolgáltatások állnak rendelkezésre, elektronikus úton (az úgynevezett IM Online rendszeren keresztül) igényelhetik például cégek nyilvános cégadatainak kivonatát, másolatát vagy mérlegadatokat, amelyek közvetlenül az Igazságügyi Minisztérium Cégnylvántartási és Céginformációs Szolgálatától származnak. Ezen egyedi adatkérésekre különböző egységárak vonatkoznak.

Amennyiben az adatszolgáltatás árazásnál megismert alapelveket akarjuk alkalmazni, azt láthatjuk, hogy a felhasználók számára az adat lekérése nem kerül semmibe, ugyanakkor a szolgáltatók különböző egyszeri belépési díjakért különböző nagyságú hozzáférést biztosítanak. Egy hatékony kétrészes tarifával állunk tehát szemben azzal az értelemben, hogy az egységár az adatszolgáltatás határkölségével egyezik meg, és a különböző nagyságú fix hozzáférési díjak szolgálják azt a célt, hogy elkülönítsék az eltérő felhasználási szokásokkal rendelkező fogyasztókat. Az alapszolgáltatás kielégíti azt az igényt, hogy az állampolgár a számára fontos pár esetben megismerhesse a szóban forgó cég alapvető adatait, és ez a korlátozott hozzáférésű szolgáltatás nulla belépési díjjal igénybe is vehető. A professzionális felhasználóknak szánt korlátlan hozzáférésű szolgáltatásért viszont magasabb belépési díjat számítanak fel, és ezen felhasználók részére további szolgáltatásokat is biztosítanak. A központilag biztosított rendszerben tehát mennyiségi, a piaci módon működő rendszerben pedig minőségi diszkriminációra látunk példát.

Bár az alapszolgáltatás ingyenes az Internet-hozzáféréssel és Ügyfélkapu-regisztrációval rendelkező ügyfeleknek, láttuk, hogy ezen „technológia” megszerzése költséges, ami jelenleg az állampolgárok jó részét kirekeszti az ingyenes adatszolgáltatásból. Az érdeklődőknek személyesen is lehetőségük van ugyan az illetékes Cégbíróságnál kikérni megyei cégek adatokat, illetve az Országos Cégnylvántartó és Céginformációs Rendszernél (OCCR) bármilyen magyarországi cég adatait. Ugyanakkor mivel ezek a dokumentumok közhitelesnek minősülnek, az áruk sokkal nagyobb az elektronikus úton hozzáférhető adatokénál: egy cég „névjegye” például 1500 forintba, míg nyilvános cégadatainak kivonata 3000 forintba kerül, és jelentős a szolgáltatás igénybevételének időigénye is. Ezen felül bár a különböző adatok reprodukálásának költségei valószínűleg közel azonosak, azok árában mégis nagy különbségek láthatóak, az ennek háttérében meghúzódó árazási logika meglehetősen esetlegesnek tűnik.

Míg tehát az online hozzáférést elemezve elsőre talán azt is mondhatnánk, hogy egy hatékony rendszerrel állunk szemben, hozzá kell tennünk, hogy az állam nem biztosít ingyenes helyettesítési lehetőséget az ezen adatokhoz való hozzáférésre. Ráadásul az az érdekes helyzet is fennáll, hogy egy piaci szereplő az állam által nyújtott szolgáltatást egyszerre olcsóbban és jobb minőségben (szélesebb hozzáférés, extra szolgáltatások) nyújtja a professzionális felhasználóknak.

6.2.2.2.A gépkocsi-nyilvántartás példája

Használt gépkocsi vásárlása során majdnem minden vásárlóban felmerül az a kérdés, hogy a megvenni kívánt jármű legális eredetű-e, nem áll-e körözés alatt. Ezt az információt a BM Központi Adatfeldolgozó, Nyilvántartó és Választási Hivatal Országos Járműnyilvántartásából, valamint a Körözési Alapnyilvántartásnak a járművekre vonatkozó részeiből lehet megtudni Interneten keresztül és telefonon. Személyes ügyintézésre nincs lehetőség.

Az egyik út a magyarorszag.hu portál Gépjárműkereső szolgáltatásán keresztül vezet, amihez a már említett Ügyfélkapu regisztráció szükséges. Amennyiben a szükséges adatok alapján a jármű megtalálható a nyilvántartásban, akkor az adatigénylő a következő adatokhoz jut hozzá: a jármű rendszáma, a hozzá tartozó forgalmi engedély szám, a jármű gyártmánya, típusa és színe, valamint hogy a jármű forgalomban van-e, és azonosító adatai alapján körözik-e (Igen/Nem). A szolgáltatás ingyenes, korlátlan esetszámban igénybe vehető.

A szolgáltatás elérhető mobiltelefonon keresztül is, a kapott adat megegyezik az internetes adatkérésnél megismerttel. A szolgáltatás csak két mobiltelefon-szolgáltató (Pannon GSM, T-Mobile) hálózatából érhető el, emelt díjas SMS formájában, amelynek díja lekérdezésenként 500 forint (a Pannon-GSM ügyfeleinek csak 470 forint). Ismerve az Internetes keresés lehetőségét, azt gondolhatjuk, két kiegészítő szolgáltatással állunk szemben, amelyből az első helyhez kötött, magasabb belépési költséggel, de kis határköltséggel, míg a második mobil, kisebb belépési költséggel, de nagyobb határköltséggel jár.

A két szolgáltatás üzemeltetői azonban nem nyilvánvaló, hogy ugyanígy gondolkodnak. Először is a BM Központi Adatfeldolgozó, Nyilvántartó és Választási Hivatal honlapján (www.nyilvantarto.hu) csak a telefonos szolgáltatást tüntetik fel, az ingyenes internetes adatkérés lehetőségéről nem tájékoztatnak. Ehhez hasonlóan a magyarorszag.hu honlapon csak az internetes adatkeresés található meg. A szolgáltatásról szóló metaadatok tehát nemcsak tökéletlenek, de nem segítenek abban, hogy a felhasználói szegmensek megtalálják a rájuk szabott szolgáltatásokat.

Nyugtalanító, hogy a telefonos szolgáltató hivatalosan nem akart információt adni a szolgáltatás forgalmáról, illetve a lekérdezési díj megosztásáról a Hivatal és a mobiltelefon-szolgáltató között. Az ügyfélszolgálat „privát” információi szerint azonban átlagosan csak 160 lekérdezés érkezik havonta. Amennyiben a szolgáltatás

ingyenessé válna, mintegy 80 ezer forint bevételkiesés keletkezne, amit az új fogyasztóknál keletkező hasznok valószínűleg messze meghaladnának.¹⁵³

6.2.3. Szakértői adatok

Az előző alfejezetben áttekintett adatok abban a tekintetben speciálisak, hogy alapvető formájukban is egyszerűen értelmezhetőek, megismerésük továbbgondolás, további feldolgozás nélkül csökkenti a döntés következményeit illető bizonytalanságot, végeredményben a döntés költségét. Számos esetben ugyanakkor a direkt megfigyelések túl nagy számúak vagy nehezen értelmezhetőek ahhoz, hogy a döntéshozó képes legyen vagy megérje számára azokat részletesen áttekinteni és ebből saját maga számára következtetéseket levonni.¹⁵⁴ Ilyen esetekben hasznosulásukhoz elengedhetetlen, hogy az ezekben az adatokban rejlő lényeges információt, amely maga is egy újabb adat, egy szakértő kitermelje, összefoglalja és értelmezze. A fogyasztók végül ezt a tömörített információt használják döntéseikhez.

A szakértői adatok közeli rokonságban állnak a későbbiekben tárgyalt kutatási célú adatokkal: az utóbbiak képezik a szakértői adatok alapját. A szoros kapcsolat ellenére a szakértői adatok és az elsősorban kutatási célú adatokra közvetlenül épülő származtatott adatok között különbség van, és azt az első fejezet gondolatmenete alapján két tekintetben tehetjük meg. Egyfelől a szakértői adatoknak potenciálisan minden állampolgár fogyasztója, míg a kutatási adatokból származtatott ismereteket jellemzően a kormányzati döntés-előkészítés, esetleg nagyvállalatok használják. Másrészt mivel a szakértői adatok jellemzően nem személyekre vonatkoznak, a kezelésükkel kapcsolatos költségek más természetűek, mint a kutatási célúaknál.

¹⁵³ A szolgáltatás megkezdése óta a lekérdezés díja nem változott, így lehetőség sem nyílik forgalmi adatok megbecsléséből arra, hogy hány új fogyasztó venné igénybe a szolgáltatást egy esetleges árcsökkenés esetén.

¹⁵⁴ Sok esetben az alapadatok nem is igazán számítanak, csak az azokból levont következtetések vagy inkább értelmezések.

Minthogy az alapadatokat többféleképpen és többféle minőségben lehet feldolgozni, a származtatott információ keletkezésének módja igen lényeges információt hordoz azok értelmezéséhez. Ezt az adatokra vonatkozó adatot metaadatnak nevezzük, és ennek a kiegészítő adatnak a közlése nélkül a származtatott adat nem foglalja össze tökéletesen a tömöríteni kívánt jelenséget.

A származtatott adatot saját céljukra felhasználó szereplők ugyanakkor a metaadatot nem mindig közlik, és ez számos félrevezető következtetésre adhat okot. Az adatközléssel kapcsolatos szabályozás pár esetben felismerte már ezt az igényt, például közvélemény-kutatások eredményeinek ismertetésekor közölni kell a mintavételre vonatkozó pár alapvető információt, de több más esetben ez még korántsem általános gyakorlat.

A szakértői adatok végeredményben igen/nem típusú döntés meghozatalát segítik olyan adatok feldolgozásával, amelyek jóval összetettebbek annál, hogy két értékkel reprezentálni lehessen őket. Ilyen döntési kényszer számos választási helyzetben felmerül, amikor az állampolgárnak feltétlenül szüksége van egy bizonyos szolgáltatás igénybevételére, amit több szolgáltató kínál. A döntéshozónak ilyenkor az alternatívák egymáshoz viszonyított hasznait kell hasonlítani az egyes esetekben, vagyis lényegében egy saját szempontjait tükröző rangsorra van szüksége. Például minden család szembenéz azzal a problémával, melyik iskolába járassa gyermekeit, szükség esetén melyik kórház vagy orvost válassza, melyik banknál nyissa meg bankszámláját vagy kösse le megtakarításait.

Mivel az állam különböző szervezetei (minisztériumok, Pénzügyi Szervezetek Állami Felügyelete, Fogyasztóvédelmi Főfelügyelőség) számos esetben finanszírozó (az oktatás illetve egészségügy túlnyomó részénél), illetve szabályozó (bankszektor, fogyasztóvédelem) szerepet töltenek be ezeken a területeken, szükségszerűen rendelkezésre állnak olyan információk, amelyekből ezek az összevetések megtehetőek. A különböző minőségbiztosítási rendszerek bevezetésével maguk a szolgáltató intézmények is rákényszerülnek, hogy értékeljék saját teljesítményüket,

illetve elkészítsék a más intézményekkel való összehasonlíthatóságot megkönnyítő statisztikákat.

A saját magukra vonatkozó összefoglaló és egyes esetekben másokkal összehasonlító információkat maguknak az intézményeknek is érdekükben áll eljuttatni a fogyasztókhoz azok meggyőzésének érdekében, ugyanakkor ezek az információk nyilvánvalóan elfogult forrásból származnak. Szükség van tehát további független szervezetekre, akik objektív szakértő módjára elvégzik ezeket a teljes körű összehasonlításokat.¹⁵⁵ Az adatigénylők így többféle forrásból is hozzájuthatnak származtatott adatokhoz és azok értelmezéséhez, amiről első ránézésre akár azt is gondolhatnánk, hogy növeli az információszerzés költségeit. A kialakult versenyhelyzet és kontrollálhatóság ugyanakkor erősebb ösztönzőket ad az információtömörítő szervezeteknek a minőségi munkára, így az adatigénylők is jobban bízhatnak az egyik választott információtömörítő szervezet által közölt információban.

A többfajta szempontból készített összehasonlításnak további előnye az, hogy a különböző igényű felhasználók így nagyobb eséllyel megtalálják az őket érdeklő kérdésre a választ. A készített rangsorok ugyanis akkor igazán hasznosak, ha azok a potenciális felhasználó számára lényeges szempontok szerinti sorrendet tükrözik. Minél több dimenziós tehát egy probléma, annál jobban felértékelődik tehát a sokféleség.

6.2.3.1. Az oktatási intézmények rangsorának példája

Az Oktatáskutató Intézet a nyolcvanas években kezdte összegyűjteni a magyar középiskolák tanulóinak eredményeit, amelyből évről-évre előállítják a középiskolák rangsorát. A kutatás céljai a három, fentiekben is elemzett szempontot találjuk meg: lehetővé tenni az iskolák vezetői és pedagógusai számára, hogy összehasonlítsák intézményük eredményeit a többi középiskoláéval; tájékoztassák a

¹⁵⁵ Mindehhez nekik hozzáférést kell biztosítani az alapadatokhoz is, amely problémára a kutatói adatok hasznait vizsgáló fejezetben visszatérünk.

középiskolákat választó tanulókat és szüleiket az iskolák kínálatáról, valamint az iskolát irányító szerveket az iskolák teljesítményéről.

A kutatás számos eredményességi mutató szempontjából összehasonlítja az oktatási intézményeket (a legjellemzőbbek a felsőoktatásban továbbtanulók aránya, a felvételi írásbeli dolgozatok átlaga és a jelentkezők nyelvvizsgáinak aránya, középiskolai versenyeredmények), ami a továbbtanulás lehetőségének irányait fontolgató családok érdeklődésére tarthat számot. Mindezek mellett ugyanakkor kontrollálni próbálják azokat az eltérő szociokulturális jellemzőket (középiskolák működési feltételei, a tanulók kezdeti felkészültsége és képességei), amelyek jelentős hatással bírnak az iskola eredményességére, és ez alapján vizsgálják azt is, mekkora hozzáadott értéket. Ezeket a „bemeneti adatokat” az eredményességi „kimeneti adatokkal” összevetve számszerűsítik az iskolák fejlesztő képességét (ezek az úgynevezett hozzáadott érték mutatók).

Az OKI 1997 óta majdnem minden évben egy összefoglaló kötetben (A középiskolai munka néhány mutatója 2004, ára 1400 forint) publikálja az utolsó öt évre vonatkozó kutatási eredményeket, amely a rangsorok mellett bemutatja a kutatási módszertant és szól a rangsorok alkalmazásának korlátairól is. A kutatás összefoglalása megjelenik a Közoktatás című oktatási hetilapban is, de ennek a folyóiratnak az elsődleges célközönsége a tanárok, az oktatási és fenntartó intézmények, alapvetően nem az iskolarendszer fogyasztóihoz szól.¹⁵⁶

Az állami kutatóintézet által végzett értelmező munka ugyanakkor láthatóan nem kap nagy nyilvánosságot. Bár az OKI honlapjáról (www.oki.hu) a teljes kutatás, illetve az összes iskolát tartalmazó rangsor letölthető, a témával foglalkozó írások többsége erre nem tesz utalást. A magyarorszag.hu Oktatás/Képzés szekciója például hosszan ír a különböző iskolatípusokról, az azokra kiterjedő törvényi szabályozásról, a felvételi eljárásokról, ugyanakkor egyetlen utalás sem esik arra, hogyan ismerhetnék meg az Internet-használók a választást megkönnyítő értékeléseket.

¹⁵⁶ A folyóiratban megjelenő cikkek csak nyomtatott formában érhetőek el.

A sajtóban az összes ezen témában megjelenő iskola-rangsor az OKI eredményeire alapul, és igazából a nagyközönség innen ismeri meg ezeket az eredményeket, jellemző példa erre a HVG-ben általában télen megjelenő hosszabb elemzés.¹⁵⁷ Ezek az írások ugyanakkor (nyilvánvalóan helyhiány miatt) egyes iskolatípusonként csak a legjobb 20-30 intézmény rangsorát közlik pár fontosabb mutató szerint, és itt szintén nem említődik meg, milyen módon juthatnának hozzá (ingyenesen) az érdeklődők a teljes rangsorhoz.

6.2.4. Többféle adatigényt kielégítő intézmények

Egyes adatszolgáltatók kezelésében lévő adatok egyszerre esnek a második (közérdekű tények), és a harmadik (szakértői adatok) kategóriába. Példa erre a meteorológiai vagy földtani adatok szolgáltatása: az állampolgárok és vállalkozások általában a szakértők által értelmezett adatokra kíváncsiak (milyen idő lesz holnap?, építhetem-e ide a házamat?), amelyek azonban nem állnak rendelkezésre közvetlenül. Az ilyen típusú adatok szolgáltatásának további jellemzője, hogy az adatpiac keresleti oldalán jelen vannak olyan professzionális felhasználók, akik az alapadatokra formálnak igényt, amelyet saját igényeiknek megfelelően használnak fel.

Ezen információk nagy részét az állam közvetve is felhasználja különböző beruházásainál, és mivel az adatfelvétel költségei igen jelentősek, indokoltnak látszik ezen adatok egy központi szervezet által történő kezelése. A hatalmas adatbázisok kezelése erősebb bevételkényszert is gyakorol ezen szervezetek működésére, ezért jó példát láthatunk különféle árazási alapelvek érvényesülésére is.

6.2.4.1. Az Országos Meteorológiai Szolgálat

¹⁵⁷ Középiskolák rangsora 2004-ben, HVG, 2005. január 12.

Az Országos Meteorológiai Szolgálat a levegőkörnyezettel és a meteorológiával kapcsolatos kormányzati feladatok ellátásáért felelős, Magyarország területére vonatkozó tevékenységet központi költségvetési szerv. A Szolgálat feladatait általánosságban a 12/199. (IV. 23.) KTM rendelet foglalja össze, valamint összesen 31 jogszabály ad feladatokat, fogalmaz meg elvárásokat, bár ezek többnyire nem részletesen megfogalmazott elvégzendő teendők. Elsődleges feladata a kormányzati szervek kiszolgálása (ÁNTSZ, katasztrófavédelem, légi közlekedés). A rangsorban ezután következik az állampolgárok tájékoztatása, majd az üzleti partnerek igényeinek kiszolgálása.

Mivel az adatok előállítása költséges, a rendelkezésre álló költségvetés szűkös, és az adatok előállításának technológiája ezt mérsékelt költséggel lehetővé teszi, a Szolgálat eredményeit piaci alapon is értékesíti. Működése és árazása több elemben megfelel annak a magatartásnak, amit elméleti megfontolásaink alapján optimálisnak értelmezünk, pár ponton viszont eltér attól.

6.2.4.1.1A meteorológiai adatok felhalmozódása és hasznosulása

A meteorológiai tevékenység jellemző tulajdonsága, hogy komoly adatfelvételi apparátust kell hozzá működtetni és nemzetközi együttműködés keretében képes igazán hatékonyan csak működni – az OMSZ nemzeti szolgálatként képviseli hazánkat a nemzetközi meteorológiai szervezeteknél. Mindkét tényező sejteti, hogy egy Magyarország méretű ország esetében a meteorológiai szolgálat természetes monopólium.

A Szolgálat adatgyűjtése és –feldolgozása a technika fejlődésével az elmúlt évtizedben egyre egyszerűbbé vált, a megfigyelések automatikusan, emberi felügyelet nélkül kerülnek be a központi adatgyűjtő számítógépbe, majd onnan ellenőrzés után a központi adatbázisba. Emberi munkaerőt főleg csak az adatok értékeléséhez, elemzéséhez használnak. Mindez azt jelenti, hogy a célként kitűzött tevékenység folytatása közben, elkerülhetetlenül nemcsak az értelmezett adatok jönnek létre, hanem egy olyan adatbázis, amely az OMSZ munkáján túl

potenciálisan más szervezetek munkájának támogatására is alkalmas. Az OMSZ kezelésében lévő adatvagyon jól dokumentált és áttekinthető: 2002. júniusától erre külön minőségirányítási rendszert alkalmaz.

Az adatokkal kapcsolatos költségek igen tetemesek, az OMSZ költségvetésének mintegy kétharmadát teszik ki (a 2004-es költségvetésben 443 millió forintot a 668 millió összköltségből). Az OMSZ működésének anyagi hátterét a költségvetésről szóló törvény határozza meg, de a működési feltételek jelentősebb részét az OMSZ saját üzleti és pályázati bevételeiből fedezi (a 2004-es költségvetésben a piaci bevételek és a költségvetési támogatások közel megegyező mértékűek). Ez azt sugallja, hogy az állam felismerte a keletkező adatok potenciális hasznosíthatóságát és ezt tette explicitte finanszírozási döntésében.

A szolgálat alapfeladatai közé tartozik az adatgyűjtés, azonban vannak olyan kutatások (mérések), amelyeket nem tartoznak a fent említett alapfeladatok közé, viszont a piac igényli ezeket. Ilyen esetekben a megrendelő fizeti a mérési költségeket, a Szolgálat pedig törekszik arra, hogy a mérési eredményeket megtarthassa, ill. az ő tulajdonába is kerüljenek. Ilyenek például a szélsébség-mérés adatai, Paksnál a radioaktív sugárzás méréséből származó adatok, Forma-1 nagydíjak idején a Hungaroringre adott részletes csapadék- és hőmérséklet-előrejelzés. A Szolgálat az adatértékesítés során arra törekszik, hogy ne nyers adatokat, hanem hozzáadott értékkel növelt adatot (azaz valamilyen szolgáltatást) adjon el. Az alapfinanszírozást biztosító támogatás ezekben az esetekben abban merül ki, hogy a megbízás teljesítése az alapfeladatok kiszolgálása érdekében létesített infrastruktúra által lehetséges.

6.2.4.1.2A felhalmozott adatok nyilvánossága

A központi finanszírozás ellenére a nem kormányzati szervek felé nyilvánosságra hozott adatok mennyisége, minősége és ára az aktuális vezetőség döntésétől függ. Az Intézet vezetőinek elmondása szerint 2005-ben az OMSZ adatpolitikája sokkal liberálisabb szemléletű a korábbi évekénél, mivel nagyobb hangsúlyt fektetnek a

felhasználók igényeinek kiszolgálására, igyekeznek igényeiknek eleget tenni. Az aktuális adatpolitika valóban előrelépés a korábbihoz képest, hiszen nemcsak nyilvánosságra hozták azokat az információkat, melyeket a jogszabály előír, az előző rezsim azonban nem hozott nyilvánosságra, de továbbiakat is elérhetővé tettek. A fejezet elején bemutatott egyszerű eljárással elvileg számszerűsíthetnénk is a nyilvánosságra hozatalból fakadó társadalmi hasznosságot. Az adathozzáférés költségei azonban időközben nem változtak, így nem nyílt lehetőség arra, hogy ezt megtegyük.

A nyilvánosságra hozott adatok köre igen szelektív, ha meg is egyezik az OMSZ létrehozásának eredeti, közszolgálati céljával: elsősorban a származtatott adatok elérhetőek, az azok forrásául szolgáló alapadatok jellemzően nem. Mindez nem olyan jelenség, amely minden történeti előzményt nélkülözne. A nemzetközi meteorológiai szolgálatok között a 80-as évek végéig korlátlan és kölcsönös adatcsere volt. Az üzleti igények erősödésével párhuzamosan megerősödött az a felismerés, hogy az országoként eltérő nyilvánosságra hozási politikák miatt ez jelentős bevételkiesést eredményez, például az USA meteorológia szolgálatai adataik döntő többségét nyilvánosságra hozzák az Interneten. A 90-es évek elején ennek szellemében született egy meteorológia szolgálatok közötti szerződés, ami az adatok két csoportját nevezi meg: amit kötelező a nemzetközi szereplőknek ingyen odaadni, és amit nem kötelező.

Az 1990-s évek végéig az OMSZ többször adott ki évkönyvet (egy éves késéssel), azonban az adatok közzétételének ezt a módját megszüntették, mivel ez sértette a szolgálat üzleti céljait: az évkönyv helyettesítője volt pénzért elérhető szolgáltatásoknak, tehát a partnerek az évkönyvet preferálták a szolgáltatásokkal szemben, ami bevételkiesést eredményezett.

Az intézet álláspontja szerint az adatok nyilvánosságra hozatalának a bevételkiesésen túl két további hátránya van:

- szakértő tudás hiányában az adatok könnyen félreértelmezhetőek (példa erre a Főpolgármesteri Hivatal tornádó-előrejelzése 2005 júliusában), amely

egyrészt pánikot okoz, másrészt rossz hatással van az OMSZ tevékenységével szembeni bizalomra;

- az adatok teljes körének nyilvánosságra hozása megbénítaná az adatforgalmat, az OMSZ honlapja nem tudna kiszolgálni a megnövekedett adatigényt.

Véleményünk szerint e két probléma közül az elsőt érdemes komolyan venni, hiszen a második pusztán finanszírozási kérdés.

6.2.4.1.3Értékesítés, árazás

Az OMSZ bevételeinek 2004-ben éppen felét adta a költségvetési és felét a saját piaci tevékenységből származó bevétel. Ebből is látható, hogy pusztán költségvetési forrásokból a meteorológiai szolgáltatások jelenlegi színvonala nem lenne tartható, és ez az OMSZ-t piaci viselkedésre kényszeríti. Tekintettel arra, hogy az alaptevékenységből keletkező adatbázis létrejötte és kezelése milyen nagymértékben automatizált és számítógépesített (amit az alternatív felhasználás nem tesz költségesebbé), a bevételi kényszer alkalmazása ebben az esetben racionális lépésnek tűnik.

Az OMSZ adatszolgáltatási politikájának speciális jellemzője, hogy az alapvető adatszolgáltatás a feldolgozott adatokra vonatkozik (mérések eredményei, rövidtávú előrejelzések, meteorológiai táviratok), ezek az információk az OMSZ honlapján ingyenesen hozzáférhetőek. Abban az esetben, ha a felhasználó az információt nem az OMSZ honlapjáról tölti le, akkor kezelési költséget kell fizetnie, melynek a minimális összege 13.400 Ft.

A nyers alapadatokat (például óránkénti adatok, repülésmeteorológiai táviratok, nyers radarképek, műholdas adatok) az OMSZ ugyanakkor csak térítés ellenében hozza nyilvánosságra, és az árazás folyamán számos diszkriminációs elvet érvényesít. Egyrészt megkülönbözteti a felhasználókat típusuk szerint: a végfelhasználók olcsóbban jutnak hozzá a kért adatokhoz, mint a média vagy más információszolgáltató-szervezet, mint ez például az alábbi árazási táblázatból

óránkénti adatok esetén látható (az árak ÁFÁ-t nem tartalmaznak). Kutatási, és oktatási célokra kért adatoknak ugyanakkor nincsen információs díja, ezekben az esetekben csak kezelési és távközlési díjat számolnak fel.

Végfelhasználó	Média	Szolgáltató	Szolgáltató viszonthatósági joggal
28 Ft	107 Ft	107 Ft	107 Ft

Ebben az esetben az árazási alapelveknél megismert diszkriminációs logikát látjuk visszaköszönni, teljes összhangban a szűkös költségvetés melletti gazdálkodás logikájával. Azon felhasználók irányában, amelyek kisebb árugalmasságú keresletet támasztanak (oktatás, kutatás), nem alkalmaznak ártorzítást, az üzleti felhasználóktól kívánják előteremteni a szükséges bevételeket. Utóbbiak esetében a média és az információ-szolgáltatók kereslete lényegesen rugalmatlanabb az OMSZ-től kapható információk iránt, így a nekik megszabott magasabb ár nem csökkenti lényegesen a keresletüket. A végfelhasználók számára, amilyen például egy nagyobb mezőgazdasági vállalkozás, számára az OMSZ információi, ha nem is tökéletes mértékben, de helyettesíthetőek, így keresletük rugalmasabbnak tekinthető, az irányukban alkalmazott ár tehát kisebb.

Látszik ugyanakkor, hogy az OMSZ árazása nem vesz figyelembe azt, hogy egyes információszolgáltatók, mint a média, esetleg az oktatáshoz, kutatáshoz hasonlóan pozitív externális hatást gyakorolnak a társadalom többi tagjára azzal, hogy maguk is egy közjószágot szolgáltatnak. A társadalmilag optimális árdiskriminációnak ezt a tényt is figyelembe kellene vennie, és ezen szereplők számára olcsóbbá tenni az adatokhoz való hozzáférést. A médiával szembeni árazásban a közszolgálati médiák sem élveznek kedvezményeket, ennek következményeként a Magyar Televízió például le is mondott az OMSZ számos szolgáltatásról.

Ezzel párhuzamosan a Szolgálat mennyiségi diszkriminációs elvet is érvényesít, a lekért adatok mennyisége és gyakorisága alapján, erre látunk példát a meteorológiai táviratok (METAR) árazásán. Ebben is megjelenik a különböző típusú

felhasználóknak szánt árazás logikája: az eseti felhasználóknak is biztosítják a szolgáltatást, igaz nagy egységáron, az állandó felhasználóik, akiknek valószínűleg igazán értékes a szolgáltatás, kisebb egységáron jutnak az adatokhoz.

Alkalmanként	Óráként frissítve			
	Egy napra	Egy hónapra	Egy félévre	Egy évre
800 Ft	4 800 Ft	7 200 Ft	21 600 Ft	43 200 Ft

Végül minőségi diszkrimináció is megfigyelhető a lekért adatok további feldolgozottsága szerint is, bár ebben az esetben nincsenek egyértelmű irányelvek, hanem tárgyalásos árat alkalmaznak.

6.2.5.A geológiai adatszolgáltatás

A Magyar Geológiai Szolgálat az államigazgatási reform során 1993-ban jött létre a Központi Földtani Hivatal jogutódjaként. A Szolgálat központi állami szervként ellátja az állam földtani feladatait (földtani szakhatósági munkák, ásványvagyon-nyilvántartás, a hazai geológiai és geofizikai adatok archiválása). A Szolgálaton belül működő Földtani Információs Rendszernek két fontos eleme van. Az első az Országos Földtani és Geofizikai Adattár, amely részben az MGSZ keretein belül, részben önállóan működő kutatóintézetek - a Magyar Állami Földtani Intézet (MÁFI) és az Eötvös Loránd Geotechnikai Adattár (ELGI) kezelődik; illetve az Építési Geotechnikai Adattár (ÉGA).

Bár az MGSZ által kezelt adatok különböznek az előzőekben áttekintett OMSZ kezelésében lévő adatoktól, az adatgyűjtés motivációja, költségstruktúrája igen sok hasonlóságot mutat. A továbbiakban csak röviden emelünk ki egy pár jellemzőt az adatok nyilvánosságra hozásáról, illetve azok árazásáról.

6.2.5.1.Országos Földtani és Geofizikai Adattár

Az OFGA jogszabályokban előírt alapfeladata az ország földtani-geofizikai megismerése során keletkező adatok, dokumentumok kezelése és szolgáltatása. A kezelésükben lévő adatbázisok nagyobbik része hagyományos táblázatos formájú adatbázis, de egyre több térinformatikai (térkép-alapú adatbázis) is található köztük. Az Adattár egyes adatbázisainak alapvető tulajdonságai világosan áttekinthetőek.¹⁵⁸

Az adatbázisok többsége nyílt adatokat tartalmaz, így ezeket külső felhasználók részére is szolgáltatják. Az adatszolgáltatás nyilvánosnak van deklarálva, a kért árban így elvileg csak az adatmásolás díját számolják fel. Ennek a költségnek a képzésére létezik ugyan egy árazási dokumentum, de a Hivatal munkatársainak tájékoztatása szerint ez a földtani adatok speciális volta nem értelmezhető könnyen, így nem jelen pillanatban nem elérhető. Ennek hiányában a költségek meghatározása inkább tárgyalásos alapon képződik.

Földtani térképek árazása esetében például azt látjuk, hogy az alkalmazott árak nagy mennyiség rendelése esetén nem térnek el lényegesen a reprodukálás határköltségétől (egy térképszelvény fotóminőségű nyomtatása 3 ezer forint/szelvény). Kisebb mennyiség esetében az egységár ugyan magasabb (1-5 szelvény esetében 8 ezer forint, 6-20 szelvény esetében 6 ezer), érvényesül tehát minimális mennyiségi diszkrimináció. Oktatási intézmények irányába ugyanakkor minden mennyiség esetében a legkedvezőbb egységárat érvényesítik. Az ország teljes területet lefedő 88 szelvény CD-n is elérhető, a magyarázatokkal ellátott adatbázis ára mindössze 3000 forint.

6.2.5.2.Építési Geotechnikai Adattár

Az ÉGÁ-t a Földmérő és Talajvizsgáló Vállalat (FTV) keretein belül 1978-ban hozták létre állami nagyberuházások, építkezések adatigényének kiszolgálására. Ezzel az adatbázissal hatékony információcserélő mechanizmust próbáltak megvalósítani: a szakintézményeknek kötelező volt az ÉGÁ-t használni, és azt szakvéleményeikkel

¹⁵⁸ Ld. www.mgsz.hu/magyar/adatbaz/index.html

gyarapítani. A FTV privatizációjakor szakmai körök nyomására az ÉGA állami tulajdonban maradt, és ekkor került az MGSZ kezelésébe.

Az adatbázis létrehozásának elsődlegesen deklarált célja a költségmegtakarítás, ami abból ered, hogy a már rendelkezésre álló adatok nem kell magas költségen „újratermelni”, vagyis az adott területen már nem kell újabb talajfeltárást végezni, vagy jóval kevesebbet. Ezek a geotechnikai adatok az építőipari közvetlen felhasználások mellett hasznos információt jelentenek például tenderek kiírásához és elbírálásához, víz- és talajminőség-változások nyomon követéséhez, majd ezek alapján környezetvédelmi tanulmányok kidolgozásához, kárfelmérések végzéséhez.

Az ÉGA honlapján explicit módon kijelenti, hogy adatszolgáltatási kötelezettségüknek forráshiány miatt nem tudnak eleget tenni, valamint bár több évtizedre visszamenően rendelkeznek adatokkal, nincs lehetőségük azok aktualizálására. Adatbázisuk fejlesztését az adatigénylőkkel folytatott információcserék révén végzik, amelyért cserébe kedvezményes díjakat határoznak meg szolgáltatásaikért.

Árazási szempontból elmondható, hogy már magának az ÉGA katalógusának a használata sem ingyenes, illetve felszámolják a szolgáltatásokhoz igénybevett munkaórák díját is. A másolási díjakon felül minél részletesebb (már meglévő) szakvéleményt kér az adatigénylő, annál magasabb az egységár (tájékoztató adatok 8 ezer forint, kis volumenű szakvélemény 15 ezer forint, közepes volumen esetén 30ezer, nagyobb volumen esetén 35 ezer forint).¹⁵⁹

¹⁵⁹ A részletes árlista megtekinthető a www.mgsz.hu/magyar/infokozp/ega/ega5.html címen.

7. A kutatási célú adatokhoz való hozzáférés hasznai és költségei

Az első fejezet elválasztotta a klasszikus közérdekű adatokat a kutatási célúaktól, a harmadik pedig bemutatta az előbbieket felhasználási lehetőségeit. Most egy jellegében más adatkört, a kutatási célú adatokat tekintjük át hasonlóképpen. Emlékezzünk vissza: a kutatási célú adatok legfőbb jellemzője az volt, hogy belőlük nem lehet közvetlenül következtetést levonni: az információ keletkezésétől fogva alakítást és értelmezést igényel. A hosszabb „termelési folyamat” nagyobb költséget, esetleg nagyobb hasznot, de mindenképpen a két tényező szélesebb skáláját valószínűsíti.

Az alábbiakban látni fogjuk, hogy a kutatási célú adatok sajátos hasznai szinte minden esetben egyéni szintű, vagy másképpen: *mikro*adatok használatával realizálható. A mikroadatok használata nemcsak kitágítja a felhasználási lehetőségek körét, de különleges, a makroszintű adatok esetében meg sem jelenő feladatok ellátását teszi szükségessé. Ilyen feladat például a személyes adatok védelmét garantáló adatkezelési eljárások alkalmazása, amely kimondottan erre az adattípusra jellemző.

7.1. Makro-, értelmezett-, mikro- és metaadatok

A kutatási célú adatok között szép számmal találunk olyanokat, amelyek nem egyénekre, egyedekre vonatkoznak, hanem egy ország, nemzetgazdaság, területi, vagy más, nagyobb egység egészére. Ilyen adat a bruttó hazai termék, az átlagbér, vagy a gyermekek száma a képzetlen háztartásokban. Ezeket az adatokat igen sokszor látjuk viszont az adat végső formájának tekinthető elemzésekben, írásokban. Kézenfekvőnek tűnik tehát, hogy elkülönítsük egymástól a mikro- és makroszintű kutatási adatokat. Ennek egyik oka az lehet, hogy az egyéni kutatási adatok – szemben a klasszikus egyedi adatokkal, melyek rend szerint gazdasági társaságokra, ingatlanokra és hasonlókra vonatkoznak – személyes jelleggel bírhatnak, ami adatvédelmi aggályokat vethet fel.

A két aggregáltsági szint elkülönítésének szükségessége az adatok nyilvánosságra hozatala szempontjából korántsem nyilvánvaló. Az, hogy miért, a makroszintű adatok keletkezésének módjával függ össze. Mi is valójában a makroszintű adat? Hagyományosan a nemzetgazdaság egészére kiszámított statisztikákat soroljuk ide: a GDP-t és annak idősorait, az egész országra jellemző halálozási rátát, az ipar termelte hozzáadott értéket. Bármennyire is különböző kérdéseket válaszolnak meg ezek az adatok, közös jellemzőjük, hogy közvetlenül nem lehet őket megfigyelni. Az „ország jövedelme” csak mint fogalom létezik, valójában a vállalkozások, személyek közössége jövedelmének összességét fedi. Akármilyen elvont is legyen a fogalom, azt a gyakorlatban a statisztikusoknak elő kell állítani, ebben az esetben éppen az ország területén belül működő cégek által termelt jövedelem összességéként. A makroszintű adat előállítását tehát az teszi lehetővé, hogy ismerjük a megfelelő elemi adatokat a vizsgált népességre vonatkozólag, illetve tudjuk, hogy milyen statisztikát kell belőlük számolni.

Vegyük azonban észre, hogy a makroszintű adatok csak speciális esetei egy tágabb kategóriának, amit nevezhetünk *származtatott adatoknak*! Az ilyen adatok két ismérvvvel jellemezhetők, a vizsgált csoportot képző ismérvekkel, illetve azzal a statisztikával, amire egy adott csoport esetében kíváncsiak vagyunk. A csoport lehet igen nagy és egyszerű, például – a makroadatok esetében – az egész országot magában foglaló, vagy egész pontosan meghatározott és kicsi, mint a felsőfokú végzettségű, 68 éves, valaha nehéziparban dolgozó férfiak csoportja. A statisztika is lehet igen egyszerű, mint az, hogy hány adott tulajdonságú egyed létezik egy adott csoporton belül, illetve összetett, mint amilyen például az adott jellemzőkkel rendelkező egyedek jövedelme, vagy akár jövedelmének szóródása egy adott időpontban az egész országban. A feltett kérdés tehát sokféle lehet, de nem változtat azon, hogy a származtatott adatok előállítása is minden esetben megköveteli megfelelő mikroadatok meglétét és – nyilvánvalóan – a kívánt statisztika ismeretét.

A származtatott adatok előállítás minden esetben felosztható az adatfelvétel és az alakítás, prezentálás lépéseire. Mindkét lépésben meghozott döntések befolyásolják azt, hogy a felhasználó milyen következtetést képes levonni a származtatott adatból – ezt a kérdést a következő alfejezetben tárgyaljuk. Igen fontos azonban az, hogy a feldolgozás folyamata maga is a származtatott adathoz kapcsolódik, és ha rögzítik, szintén adatként értelmezhető. Egy eredmény felhasználásához nem csak annak tartalmi leírása, de annak ismerete is fontos, hogy milyen módszerrel készült, hiszen egyik vagy másik módszer alkalmazása jelentősen befolyásolhatja az eredményeket.

Az adatok feldolgozásában lehetséges különbségek érzékeltetésére tekintsünk egy példát! Az Állami Foglalkoztatási Szolgálat (ÁFSZ) „Bértarifa” felvétele gazdasági vállalkozások közül választ ki mintegy 20 ezret, amelyek dolgozóinak néhány személyes ismervét (például iskolázottság) és bérét rögzíti. Noha a „Bértarifa” felvétel nem teljes körű, az egyedi adatok sokfélesége számos fontos kérdés megvizsgálását teszi lehetővé. Tegyük most fel, hogy a 60 éves emberek átlagkeresetére vagyunk kíváncsiak! Egyszerű átlagot számolva 136.464 forint jön ki. Ez az eredmény azonban nincs tekintettel arra, hogy a mintavétel nem teljesen véletlenszerű, hanem abban különböző ágazatokból (a népességben tapasztalhatóhoz képest) különböző arányban kerülnek be vállalkozások. Ha az adatok *súlyozásával* a minta ezen jellegzetességét figyelembe vesszük, 128.494 forintot, mintegy 8 ezer forinttal kevesebbet kapunk, mint korábban. A különbség 6 százalék, ami nem rettenetesen nagy, de ilyen nagyságú különbségekről parázs vita szokott zajlani az érdekegyeztetésben, és a családok költségvetésében is különbséget jelenthet.

Mi a feltétele annak, hogy ezt a különbséget figyelembe vehessük? Többek között a mintavétel folyamatának pontos ismerete. Ha erre nincs lehetőség, a megfelelő súlyok nem készülhetnek el. A súlyozás azonban nem csak egy módon végezhető el. Lehetnek olyan ismérvek, amelyeket egy kutató fontosnak tart, mások pedig nem. Lehetséges az is, hogy a súlyozásba hiba csúszott. Nem teremt tehát ideális helyzetet az sem, ha egy adatgazda maga elvégzi a szükséges korrekciót, de nem

osztja meg az elkészítéshez szükséges információkat. Természetesen mindez a megfelelő változtatásokkal érvényes az adatok feldolgozására, a származtatott adatok előállításának módjára.

7.2. Statisztika és hatásvizsgálat

A származtatott adatok tipikus esete a statisztikai tábla, amelyet a statisztikai hivatalok létezésük kezdete óta gyártanak. Ez a tábla a származtatott adatok bemutatásának tipikus formája. Példaként álljon itt a Magyar Statisztikai Évkönyv 2003-as kiadásából egy táblázat (a 4.15-ös táblázatnak csak a férfiakra vonatkozó első része):

Az alkalmazásban állók havi bruttó és nettó átlagkeresete*, 1992–2003

Év	Bruttó átlagkereset			Nettó átlagkereset		
	fizikai foglalkozásúak	szellemi	együtt	fizikai foglalkozásúak	szellemi	együtt
1992	19 160	39 966	24 505	13 991	25 200	16 870
1993	23 321	48 455	30 106	16 419	29 604	19 979
1994	28 619	60 395	37 509	20 632	37 675	25 400
1995	33 528	70 334	44 087	23 469	42 159	28 831
1996	40 168	84 408	52 879	27 563	49 227	33 787
1997	48 369	106 786	65 186	33 502	64 391	42 395
1998	52 421	118 914	71 931	37 014	72 128	47 317
1999	60 627	141 833	84 649	41 230	84 047	53 896
2000	68 556	162 839	96 563	45 587	95 182	60 319
2001	80 042	190 628	113 324	52 723	109 398	69 780
2002	91 384	219 039	131 413	61 813	128 565	82 745
2003	98 173	246 984	147 345	70 306	140 815	93 604

* 1992–1997-ig a 20 fő feletti, 1998-tól a 4 fő feletti létszámú gazdasági szervezetek teljes munkaidőben foglalkoztatottainak adatai.

Forrás: éves intézményi munkaügyi adatgyűjtési rendszer.

A táblázatból időben és tulajdonságok szerint (fizikai/szellemi) megbontva közöl átlagos adatokat két ismervre, a nettó és bruttó keresetre. Világosan leolvasható belőle, hogy a keresetek folyamatosan nőttek a vizsgált időszak alatt mind bruttó, mind nettó értelemben. Kiderül az is, hogy a szellemi munkát végzők átlagosan többet keresnek, mint a fizikaiak, és még az is kiolvasható, hogy az utóbbiak átlagos effektív adókulcsa magasabb, mint az előbbieké. A táblázat szerkezetével a mikroadatok kezelője, jelen esetben a KSH, meghatározta, hogy a mikroadatok által hordozott információból a következtetések milyen körét vonhatjuk le. Ez a statisztikai munka klasszikus feladata: a végtelen változatosságú adatokat oly

módon sűríteni, hogy a felhasználó könnyedén választ kapjon az őt érdeklő kérdésekre.

A statisztikai megközelítés tökéletesen alkalmas egy adott helyzet leírására: megtudjuk, hogy egy adott helyen és pillanatban mi mennyi. Sokszor azonban nem csak erre vagyunk kíváncsiak, hanem arra is, hogy ez vagy az *miért* annyi, amennyi. Ha történik valamilyen esemény, szeretnénk megismerni annak a hatását. Hogy közelebb kerüljünk a kérdéshez, térjünk vissza a fenti táblázathoz. Tudjuk, hogy 2001-ben és 2002-ben jelentősen megemelték minimálbér összegét. A politika célja szerint a lépéssel ösztönözni szeretne volna a legális munkavállalást a bérek, mindenekelőtt a legalacsonyabb növelésével. Vajon a fenti táblázat alapján meg tudjuk mondani, hogy mi volt az intézkedés hatása a bérekre, különösen a keveset keresők bérére?

A táblázat alapján az átlagos bérek változását követhetjük. Ahhoz, hogy kérdésünket meg tudjuk válaszolni, tudnunk kellene, hogy mi történt volna a bérekkel akkor, ha a minimálbér nem emelkedett volna. Noha ezt természetesen nem tudhatjuk meg soha, következtetni tudunk rá, mégpedig a bérek változásának mechanizmusát követve. Tudjuk, hogy a bérek alakulása a piaci erők interakciójának, a bérek és a foglalkoztatás differenciált igazodásának eredménye, így számtalan kérdéssel kell szembenéznünk. Azért nőttek a bérek, mert emelkedett a minimálbér, vagy azért, mert egyébként is fellendülést él át a gazdaság? A munkavállalók milyen hányadának a bére van közel a minimálbérhez – mekkorát „harap” az emelés? Azért nőttek a bérek, mert a munkaadók el tudták „nyelni” a költségkülönbséget, és ugyanazokat az embereket foglalkoztatják, vagy azért nőttek, mert erre nem voltak képesek, és meg kellett válniuk sok alkalmazottjuktól, ami emelte az átlagbért? Csak a korábbi minimálbéren keresők bére emelkedett, vagy másoké is? A minimálbértől milyen „távolságra” nőttek a bérek, hogyan változott a bérek eloszlása?

Egy adott problémával kapcsolatban a lehetséges kérdések és így az alkalmazható eljárások száma is végtelen. Még ha kialakulnak is rutinná váló számítások, a

tudományos kutatás egyre újabb és újabb módszereket alakít ki - statisztikai hivatal legyen a talpán az, aki képes mindet megelőlegezni, elébe menni az igényeknek, és a tudományos közösséget a szükséges számításokkal proaktív módon kiszolgálni. Különösen igaz ez akkor, ha belátjuk, hogy a statisztikai tábla nem az egyetlen adatközlési forma, mint ahogyan az átlagos érték (kiemelt szerepe ellenére) nem az egyetlen statisztika, amely érdeklődésre tarthat számot. A fenti példában már említettük, hogy a minimálbér emelésének hatását csak akkor tudjuk számszerűsíteni, ha meg tudjuk mondani, mennyivel nőtt volna a vizsgált népesség bére a minimálbér emelése nélkül is. Erre a kérdésre az egyszerű átlagnál jóval összetettebb statisztika képes csak válaszolni.

A tanulság azonban nem korlátozódik példánkra: a hatásvizsgálatok esetében kivétel nélkül alapvető fontosságú, hogy meg tudjuk mondani, mi lett volna akkor, ha nem történik meg a vizsgált beavatkozás. Ez pedig vagy egy adatfelvétel gondos, a hatásvizsgálat logikáját felismerő és alkalmazó megtervezésével lehetséges vagy olyan, statisztikai alapú technikákkal, amelyek egytől egyig rendkívül finom, egyéni adatokat igényelnek.

7.3. Az állam hatékony működtetése – új igények

Az, hogy a statisztikai hivatalok táblázatokot gyártanak, nem a véletlen műve. Amellett, hogy a táblázatos forma sok esetben igen informatív, a mai számítástechnikai fejlettség előtt az adatok közlésének egyetlen formája volt (a táblázatokkal analóg grafikonok mellett). A helyzet mára radikálisan megváltozott: a technika és az alkalmazott statisztika fejlődése lehetővé tette, hogy íróasztali számítógépekkel rendkívüli méretű adatbázisokat kezeljünk, és olyan matematikai-statisztikai számításokat végezzünk, amelyet 30 évvel ezelőtt csak az akkori szuperszámítógépek voltak képesek.

Nemcsak a technikai háttér, vele együtt a felhasználási igények is fejlődtek. A változások hatásának vizsgálata, melyre fent már bemutattunk egy példát, számtalan esetben érdekes gyakorlat lehet, a hatékony gazdálkodás azonban

egyenesen megköveteli, hogy a meghozott döntések eredményét mérni tudjuk. A probléma összetettsége miatt erre hosszú ideig nem volt mód, mára azonban a programértékelés rutinná vált. A kérdés fontosságát mi sem mutatja jobban, mint a programértékelés terén született iszonytató mennyiségű tanulmány, illetve az, hogy a 2000-es közgazdasági Nobel emlékdíjat James Heckman, a programértékelés módszertanának központi figurája kapta.

Az angolszász országokban, különösen az Egyesült Államokban a különféle programok hatásvizsgálata bevett gyakorlat, míg Európában csak az utóbbi időben jelenik meg hangsúlyosan. A versenyképesség fokozása, az államháztartási hiány és az adóteher csökkentésének együttes igénye olyan ambiciózus célkitűzések, amelyek az általánosnál még fontosabbá teszik annak vizsgálatát, hogy az adófizetők pénzét mire költi a kormányzat. Magyarországon a programértékelés iránti igény még mérsékelt, de mivel az Európai Unióból érkező támogatások jelentős része esetében megkövetelik a programok szigorú értékelését, ez a helyzet várhatóan meg fog változni. Nem lehet eléggé hangsúlyozni: ezek az értékelések csak és kizárólag mikroadatokon végezhetők el.

7.4. A mikroadatok előállításának és felhasználásának „technológiája”

Ahhoz, hogy rámutassunk, milyen hatással lehet a kutatási célú adatok hozzáféréseinek lehetővé tétele, látnunk kell, hogy milyen termelési folyamat zajlik le addig, míg a nyers adatokból végtermék válik.

Ahogy arra korábban utaltunk már, a mikroadatok két forrásból származhatnak: vagy adminisztratív forrásból, vagy külön erre a célra szervezett adatfelvételtől. Az előbbire példa a lakcímnnyilvántartás, az OEP által fenntartott gyógyszerfogyasztás és orvosi ellátások igénybevételeinek adatbázisa, az ONYF nyugdíjadatbázisa, a munkavállalók munkatörténetét nyilvántartó NYENYI adatbázis, amelynek helyét, ha lassan is, de valamikor feltehetőleg átveszi az Egységes Magyar Munkaügyi Adatbázis (EMMA). A célzott adatfelvételekre példa a KSH által készített

Népszámlálás, a Háztartási Költségvetési Felvétel vagy az Állami Foglalkoztatási Szolgálat „Bértarifa” felvétele.

Az egyes adatbázisok létrejöttekor igen fontos *metaadat* jön létre, amely éppen az adat születésének folyamatát írja le. Ide tartozik minden olyan leírás, amely dokumentálja az információ szerzésének módját, a felhasznált kérdőíveket, a kérdezők felkészítését, a számukra kiadott utasítást, az adatok kódolásának menetét, egy szóval minden olyan mozzanatot, amely az adatok keletkezésével összefügg. Mindez természetesen nem öncélú: az adatokból levont következtetést nemcsak ezek, hanem még olyan, látszólag lényegtelen momentumok is befolyásolhatják, mint a kérdező életkora. Ezeknek az adatoknak a nyilvánossága az alapadatokhoz hasonló jelentőségű.

Adminisztratív adatok esetében az adatbázisok minden időben személyes adatokat tartalmaznak, hiszen az egyének azonosíthatósága a munkavégzés feltétele. A célzott felvételek során keletkezett adatok esetében a személyes azonosítóktól már az adatok digitalizálása után megfosztják azokat, a végső adat személyes jellegét ezekben az esetekben azok részletessége eredményezheti. Ha ugyanis egy adott területi egységen belül egy felvétel tartalmaz adatokat egy igen atipikus egyedről (például egy falusi orvostól), az tulajdonságai alapján egyértelműen beazonosítható. Szélsőséges esetben tehát az azonosító törlése önmagában nem megoldás, mert a tartalmi adat maga működik akként. A KSH által kiadott adatok azonban minden esetben meg vannak fosztva a részletes területi azonosítóktól, kistérségi szinten pedig egyéni jellemzők alapján lehetetlen bárkit is beazonosítani.

Maguk az egyes egyénekhez tartozó adatsorok (rekordok) kutatási célú felhasználás esetében utoljára az adatok javításakor, tisztításakor játszanak szerepet, például ha időben többször kérdezik meg szándék szerint ugyanazt a háztartást, de az egymást követő felvételekben alapvető adatok, mint a születési év és a nem, nem egyeznek meg. A tisztítást végző személy itt közvetlenül tekinti az egyedi adatokat, mindez azonban általánosságban adatvédelmi szempontból semmilyen problémát nem vet fel, hiszen az adatok anonimizáltak.

Hasonlóan az adatok keletkezésének dokumentálásához, az adatok javítási módjának pontos ismerete is rendkívül fontos. A tisztítás során nem ritkák az elkerülhetetlenül önkényes döntések, melyeknél ismét fontos tudni, hogy milyen döntés született. Igen fontos példa erre a hiányzó adat javítása, pótlása. Tegyük fel, hogy ismerjük a háztartások jövedelmét és az azokat csökkentő tételeket (például adó, TB-járulék), és azt vesszük észre, hogy néhány háztartás esetében a nettó jövedelmek negatívak. A negatív jövedelmek nem kevés elemzés esetében gondot okozhatnak, így valószínű, hogy a helyzet nem tartható. Lehetséges megoldás az, hogy a negatív jövedelmeket 0-val helyettesítjük. Ha azonban az eredeti problémát a nettó jövedelem valamely komponenséhez adott szimmetrikus, tehát pozitív és negatív irányba átlagosan ugyanannyit tévedő hiba okozta (olyan, amely mondjuk jellemzően 0, de -1 és 1 között bármi lehet), akkor ezzel a változtatással aszimmetriát hoztunk létre, aminek kedvezőtlen hatása lehet bizonyos elemzésekre. A helyes eljárás az, ha az ilyen változtatásokat dokumentálják, és a változtatás nélküli adatsor is elérhető marad.

Kiemelt figyelmet érdemel az imputálás, az adatjavítás speciális esete. Ennek során valamilyen megfontolás alapján (például a „hasonló” egyedek meglévő adatai alapján) generálnak hiányzó adatokat, amelyeket beillesztenek az eredeti, valódi adatok közé.¹⁶⁰ Az imputálás szélsőséges, de a gyakorlatban alkalmazott esete az, amikor nem létező adatrekordokat hoznak létre valamilyen, egzogén forrásból ismert arányosság elérése érdekében. Nyilvánvaló, hogy az eljárás dokumentálása milyen jelentőségű ebben az esetben, hiszen ha az adat felhasználójának nincs tudomása az imputálás tényéről, nem tudja, minek köszönheti a kapott eredményt: az adatkezelő megfontolásainak, vagy valódi összefüggésnek.

Az adattisztítás sajátos szempontot vet fel akkor, ha a mikroadatokat nem kifejezetten mikroelemzés támogatására készítették el. Amint arról már többször szó esett, a statisztikai hivatalok eredendően nem mikroelemzések céljára, hanem statisztikai táblázatok, aggregált kimutatások elkészítéséhez vették fel a

¹⁶⁰ Amennyiben egy mintában 100 egyetemi hallgató van, de kettőnek nem ismert a jövedelme, akkor ezt a két hiányzó adatot helyettesíthetjük például a 98 megfigyelés átlagával.

mikroadatokat. Az olyan esetekben, ahol ez a megközelítés változatlan, az egyedi adatok minőségellenőrzése kevésbé feszes. Ennek éppen statisztikai oka van: ha viszonylag nagy aggregátumokra számolunk átlagot kellően sok egyedi adatból, a véletlen hiba éppen statisztikai tulajdonságai következtében elhanyagolható. A hagyományos feladatkört ellátó statisztikai hivatal így teljesen racionálisan nem alkalmaz olyan részletekbe menő ellenőrzést és javítást, ami a mikroelemzések esetében elengedhetetlen, tekintve, hogy az ott megcélzott eredményeket már lényegesebben befolyásolják a felderítetlen hibák.

Ha az adatbázis „rendben van”, a kutatás különféle irányokat vehet. Az egyik felhasználási irány az, hogy a kutató a statisztikai hivatal tábláihoz hasonló, de azokkal nem megegyező táblázatokat készít (ilyen táblázatokra bőségesen találunk példát a Munkaerőpiaci Tükör című kiadvány statisztikai blokkjában). A másik, jóval összetettebb felhasználási módszer az, ha a kutató összefüggéseket próbál kimutatni két vagy több adat között, bizonyos körülmények rögzítésével. Anélkül, hogy ennek részleteibe belemennénk, két megjegyzést tennénk. Az első, hogy éppen ez a vizsgálati módszer az, ami a hatásvizsgálatok elvégzéséhez szükséges, tehát az adatok közjósággént való hasznosulásához elengedhetetlen. A második, hogy míg az adatok tisztítását, illetve összetett táblázatok elkészítését megfelelő anyagi háttér mellett bármely intézmény vállalhatná, az adatok ilyen összetett vizsgálatát csak az a kutató tudja elvégezni, aki a vizsgált kérdést megfogalmazta, akinek a kutatás előrehaladtával eszébe jutnak ötletek és adott esetben új vagy újszerű módszereket is ki tud dolgozni. Mivel a kutatás részben exploratív feladat, a kutató igen sok módszert kipróbál, mire eljut ahhoz, amellyel végül a leghatékonyabban rá tud világítani az általa vizsgált probléma lényegére. A kutatási célú adatok felhasználásának utolsó fázisa az, hogy a kutatás eredményeit jellemzően folyóiratcikk formájában publikálják. Ez a kutató lehetősége arra, hogy ne csak eredményeit, hanem a felhasznált módszereket is részletesen dokumentálja. A lehetőség ma már kényszer: magára bármit is adó folyóirat megköveteli az eljárások olyan részletességű leírását, hogy az érdeklődő rekonstruálhassa az alkalmazott eljárást.

7.5. Mikroadatok használata a gyakorlatban: hazai tapasztalatok a mikroadatok hasznosításával kapcsolatban

Az eddigiekben főleg elméleti megfontolások alapján érveltünk amellett, hogy a mikroadatok hozzáférhetősége igen jelentős társadalmi haszonnal járhat. Annak ellenére, hogy számos mikroadat jelenleg nem hozzáférhető, másokkal már ma is dolgozhatnak kutatók. Az alábbiakban néhány ismertebb, már lezajlott, az állam működését alapvetően érintő kutatások eredményeit mutatjuk be röviden. Igyekszünk rámutatni arra, hogy kizárólag makroszintű adatok ismeretéhez képest milyen tudásbeli előrelépést tesz lehetővé az, hogy e tanulmányok mikroszintű adatokat használtak. Végül röviden szólunk a kutatási célú adatok nyilvánosságának egy sokkal közvetlenebb hasznáról is, amely az adatgazda és a kutató közötti kétirányú kommunikáció következtében alakul ki. Ez a kommunikáció az adatgazdák és a kutatók között a világon mindenütt jól bevált munkamegosztást alakított ki: az előbbiek kiváló minőségű adatokat, az utóbbiak ez alapján hasonló elemzéseket állítanak elő.

7.5.1. Összetételhatások kiszűrése mikroadatok segítségével: a cigányság foglalkoztatása¹⁶¹

A cigányság foglalkoztatása a szociál- és munkaügyi politika egyik legfontosabb kérdése. Magyarországon 1993-ban a munkaképes korú cigány és nem cigány népességé aktivitása között 26 százalékpont volt a különbség, és a helyzet azóta sem javult. A cigányság igen jelentős része ma segélyeken és közmunkán él, amiből a „normál” foglalkoztatásba való átlépés esélye minimális. Mindez reménytelen helyzetet teremt az érintett cigányok számára, és egyben óriási terhet ró a költségvetésre. A kérdés tehát valóban rendkívüli fontosságú.

Korábban a statisztika és hatásvizsgálat közötti különbségekről szólva egy táblázat példájával illusztráltuk, hogy a mikroadatok hozzáférhetősége a lehetőségeknek milyen tárházát nyitja meg a felhasználó előtt, aki rugalmasan akar statisztikai

¹⁶¹ Az itt leírt összefoglalást Kertesi Gábor (2000) motiválta, az adatok forrása a tanulmány 11. táblázata.

táblákat készíteni, azokon keresztül kutatási kérdéseket megvizsgálni. A táblázatban kimondva, kimondatlanul, egy tényező hatását vizsgáltuk egy másikra: példánkban az egyik az idő, illetve a foglalkozás típusa (fizikai/szellemi) volt, a másik a bér. Tekintsük most a cigány foglalkoztatás példáját! Legyen az a hatás, amit meg szeretnénk érteni, a 40-54 éves cigány férfiak foglalkoztatási esélye (100 cigányból hány dolgozik) az iskolai végzettség függvényében – számos tanulmány igazolta, hogy a foglalkoztatás esélyének egyik legjobb előrejelzője az iskolázottság. Tekintsünk csak két lehetőséget, a nyolc általános iskolai, és a szakmunkás végzettséget. A kérdés rendkívül egyszerű táblázatos formában a következőképpen írható fel (a táblázat celláiban a foglalkoztatási hányadok, zárójelben a hozzájuk tartozó adott népességbeli arányok találhatók):

	8 általános	Szakmunkás
Nem cigány	63% (50%)	75% (50%)
Cigány	33% (76%)	51% (24%)

A fenti táblázat származtatott adatnak tekinthető. Ha a kutató pusztán ezt ismeri, rámutathat arra a rendkívül fontos tényre, hogy a cigányság alacsony foglalkoztatása nemcsak a képzett cigány emberek kis arányának tudható be, de annak is, hogy esetükben a foglalkoztatás a képzettség nagy „hozadéka” ellenére jóval kisebb, mint a nem cigányoknál: a szakmunkás végzettségű cigányok foglalkoztatása 50 százalékkal alacsonyabb, mint a nem cigány szakmunkásoké. A különbség első látásra óriási diszkriminációra utal, aminek messzemenő következményei vannak a megfelelő szakpolitikák teendőire vonatkozólag.

Kérdés azonban, hogy csak a diszkrimináció következményeit látjuk-e a különbségekben? A közgazdaságtudomány eredményeiből ismert, hogy a foglalkoztatási esélyre az iskolázottságon kívül más tényezők is hatnak, többek között a gyermekek száma, a családi állapot, a szülők végzettsége (a végzettségben nem megjelenő tudás közelítéseként) és a munkavégzéstől független jövedelmek megléte/nagysága. A fenti táblázat ugyan bonyolítható lenne még egy ismérvvvel, de többel nem, így ezeknek a tulajdonságoknak a hatása együttesen nem választható el az iskolázottság hatásától. Ez persze önmagában nem lenne baj, ha a felsorolt

ismérvek egymástól függetlenek lennének. Csakhogy nem azok: az iskolázottaknak átlagosan kevesebb gyerekük van, kevésbé kapnak segítyt (részben persze azért, mert dolgoznak), „jobb gyerekszobában” nevelkedtek, és sorolhatnánk még az ismérvek maguk közötti összefüggéseket. Ezeket a jellemzőket egytől egyig figyelembe tudjuk venni akkor, ha egyéni adatokkal dolgozhatunk és megfelelő statisztikai eljárást alkalmazunk.

Minthogy a fenti tényezők mind az iskolázottsággal összefüggésben magyarázzák a foglalkoztatási esély különféle típusú emberek közötti különbségeit, árnyalják az egyszerű eszközökkel kimutatott különbséget. A 1993-1994-es cigányvizsgálat adataival megtöltve az elemzési keretet rámutathatunk, hogy bár a most felsorolt „háttérjellemzők” hatása igen jelentős, a cigány és nem cigány népesség között továbbra is számottevő eltérés mutatkozik abban, hogy egy iskolai fokozatnyi előrelépés milyen hatással van a foglalkoztatási esélyekre. Nem arról van tehát szó, hogy a cigányság foglalkoztatása azért alacsony, mert átlagosan nagyobb családokban, több gyerekkel, több segélyben részesülő családdal élnek együtt.

7.5.2. Az összetételhatás és az időbeliség kölcsönhatása: a munkanélküliségből kilépés valószínűsége és a jövedelmi mobilitás

Mikroadatokat segítségével nemcsak pontosabban vizsgálhatunk egyes, makroadatokból már ismert jelenségeket, de olyanokat is megmutathatunk, amelyek kizárólag mikroszintű adatokból konstruálhatók: ilyen a munkanélküliségből kilépés folyamata. Ezt a folyamatot az úgynevezett túlélési függvény írja le, amely azt mutatja meg, hogyan alakul a munkanélküliségből kilépés esélye a munkanélküliségben eltöltött idő függvényében. Bár mint láttuk, gyakorlatilag minden makromutató mikroadatokra épül, a származtatott statisztika az esetek többségében igen egyszerű: a számítás során az egyedek szerepe felcserélhető, a származtatott adathoz való hozzájárulásuk marginális. A túlélési görbe esetében ez nem így van, mert minden résztvevő adat reprezentál egy, a

munkanélküliségben eltöltött időszakot, és majdnem önállóan, a hozzá leghasonlóbb adatpontokkal együtt szerepel.

A munkaügyi politika egyik központi célja a munkanélküliség leszorítása, az érdeklődés középpontjába így igen sokszor kerül a munkanélküliségi ráta. Ez a mutatószám azonban statikus, nem tájékoztat arról, hogy miként alakul a munkanélküli állomány összetétele. Munkanélküli ellátásra 2003-ban 67 ezer ember volt jogosult. Politikai szempontból korántsem mindegy, hogy túlnyomó részben új belépők adják a megfigyelt állományt, akik rövid időn belül ki is lépnek onnan, vagy létezik egy jelentős méretű kemény mag, amely nem képes kiszakadni a munkanélküliségből. Az arányoknak a munkanélküli ellátás rendszerére vonatkozóan messzemenő következményei vannak. A munkanélküliségből könnyen, „hatékonyan” kilépni tudó csoportnak nagyobb összegű, rövid ideig tartó anyagi segítségre, míg azoknak, akik valóban képtelenek kitörni a munkanélküliségből, inkább szociális ellátásra van szüksége.

A munkanélküli ráta mikroadatokból standard módszerrel számított mutató. Ugyanezen mikroadatokból kiszámítható és ábrázolható a munkanélküliségből kilépés esélye is az eltelt idő függvényében. Ilyen összefüggés bemutatásával vizsgálja például Bódis-Galasi-Micklewright-Nagy (2005) 5. ábrája a munkanélküli járadék jogosultsági kritériumának 1993-as szigorítását az ÁFSZ munkanélküli regiszterének mikroadatai alapján. Az ábrából azonban nem tudjuk meg, hogyan alakul e folyamat a különféle csoportok, a nők, képzettek, adott régióban élők, stb. esetében. Erre, az előző példához hasonlóan, ismét statisztikai becslésekkel kaphatunk választ. A különbséget a számítás részletein túl az jelenti, hogy azok elvégzéséhez nemcsak hogy mikroadatokra, de ráadásul úgynevezett panel adatokra van szükség, amelyek egy-egy ember időbeni követését teszi lehetővé.

A munkanélküliséggel analóg probléma a jövedelmi egyenlőtlenségeké. Ma gyakran elhangzó érv: az igazán leszakadókon kell segíteni, ez kell, hogy a politika célja legyen. De vajon hányan vannak ebben a csoportban? Mennyi ideig szegény a

szegény? A munkanélküliséghez hasonlóan a szegénység kezelésére is más és más módszer alkalmazandó akkor, ha a kérdésre ez vagy az a válasz.

A szegénység, a jövedelmi helyzet esetében is konstruálhatók különféle mutatók, amelyek azt mutatják, hogy különféle határokat meghúzva a népesség milyen aránya található az adott határ alatt. Az a mutató ismét nem informatív arra nézve, hogy milyen sebességű a népesség mozgása a „szegény” és „nem szegény” státusok között. A munkanélküliséggel szemben azonban a szegénység fennállása nem dönthető el igen/nem kérdéssel – a „szegénység” címkéje minden esetben önkényes. Mikroadatokat segítségével nemcsak elkerülhető ez az önkényesség, de minőségileg is túl lehet lépni a szegény/nem szegény dichotómián, hiszen rugalmasabb, árnyaltabb eszközöket használhatunk ugyanannak a gondolatnak a megjelenítésére. Lehetőség nyílik például arra, hogy megvizsgáljuk, az egyes jövedelmi szinteken (ötödök, tizedek, vagy bármilyen más beosztás szerint) élő emberek jövedelmi helyzete milyen irányban és mértékben változott az idők folyamán. Ilyen számítást reprodukál az alábbi táblázat a KSH Háztartási Költségvetési Felvételéből származtatott „Rotációs Panel” segítségével:

Az egy fogyasztási egységre eső jövedelem 10 százalékos szintű mobilitása a jövedelmi decilisek százalékában 1996-2001 között, %

decilis	1996-1997		1997-1998		1998-1999		1999-2000		2000-2001	
	Le	Fel	Le	Fel	Le	Fel	Le	Fel	Le	Fel
1	0	40	0	37	0	32	0	34	0	22
2	4	35	5	36	7	39	8	34	9	35
3	24	31	15	44	21	29	16	32	14	32
4	24	39	25	38	34	22	23	32	23	38
5	24	36	28	34	27	29	31	27	30	30
6	29	32	36	25	32	21	35	20	29	28
7	41	22	33	26	36	22	31	21	37	25
8	34	10	40	16	32	16	37	14	41	8
9	44	6	34	7	36	4	32	3	24	3
10	35	0	19	0	16	0	19	0	18	0
Átlag	26	25	24	26	24	21	23	22	22	22
Összes	51		50		45		45		44	

Forrás: Kapitány – Molnár – Virág . 16. táblázat. A szerzők saját számítása az 1995-1998, az 1998-2000 (az 1998-1999-es átmenetre vonatkozóan) és az 1999-2001 évi rotációs panelek alapján.

A mobilitást bemutató táblázat alapján világossá válik: bár a 90-es évekhez képest mérséklődött a folyamat, a mai napig igaz, hogy az egyenlőtlenségek növekednek. A már eleve jómódúak azok, akiknek a helyzete tovább javul, és a szegények azok, akiké tovább romlik. Ez az eredmény összhangban van sokunk mindennapi tapasztalataival, de jóval megalapozottabb annál, így a szociálpolitika bátran építhet rá. Bemutatását csak az tette lehetővé, hogy létezik olyan lakossági adatfelvétel, amely követhető módon szolgáltat részletes információt a háztartások jövedelmi viszonyairól egymást követő években.

7.5.3. Döntéshozók viselkedését jellemző modellek, paraméterek becslése mikroadatokból – a minimálbér emelésének vizsgálata fiskális szemszögből

Mikroadatok segítségével nem csak mikroszintű elemzések végezhetők. A makroszintű, az egész gazdaság működésének leírását célzó modellek valaha ad hoc feltételezéseken alapultak, ma azonban már szinte kivétel nélkül alulról építkeznek, az alkalmazott viselkedési szabályoknak mikroszintű megalapozottsága van. A vállalatok racionálisan gazdálkodnak a munkaerővel és tőkéjükkel, és a fogyasztók sem találomra vagy hüvelykujj-szabály szerint költik el rendelkezésre álló pénzüket. A mikroadatok segítségével tehát realitással tölthetjük meg az elméleti konstrukciókat. A modelleket egy adott gazdaság működésére jellemző viselkedésűvé szabhatjuk, és az így előálló modellel a gazdaságpolitika számára releváns válaszokat adhatunk.

2001-ben a kormányzat közel kétszeresére emelte a minimálbért. Az intézkedéssel kapcsolatos várakozás az volt, hogy az emelkedő bér vonzóbbá teszi a legális munkavégzést, aminek következtében majd megemelkedik annak szintje. A kormányzat ugyanakkor úgy tűnik, hogy nem számolt az emelés által kiváltott negatív folyamatokkal, amelyek a csökkenő munkakeresleten keresztül fejtik ki hatásukat. Ahhoz, hogy számot tudjunk vetni, melyik hatás volt erősebb, olyan, a

nemzetgazdaság egészére kiterjedő modellre van szükség, amely képes kell részletességgel leírni egy ilyen beavatkozás tovagyűrűző hatásait.

A fenti tulajdonságokkal rendelkező teljes modell túl bonyolult ahhoz, hogy ezen a helyen ezt áttekintsük (teljességében Halpern és szerzőtársai (2004) fejtik ki). Mi itt annak csak azt a blokkját emeljük ki, amelyik megjeleníti a vitatott kérdést: mi a minimálbér emelésének hatása a foglalkoztatásra nézve. Az emelés foglalkoztatási hatását az határozza meg, hogy a vállalatok mennyire reagálnak a munkaerő árának, a bérnek változására. Racionális gazdálkodást feltételezve a válasz egyértelmű: a vállalatok növekvő bérek mellett csökkentik a foglalkoztatott munkaerő létszámát. Azt, hogy egy százaléknyi béremelés a munkakeresletnek milyen mértékű változását (csökkenését) vonja maga után, a munkakereslet úgynevezett bérrugalmassága mutatja meg.

Bár az elméleti válasz egyértelmű, ez még nem jelenti azt, hogy ezt a tapasztalati tények is visszaigazolják. Noha az ehhez szükséges mérés elvileg makroszinten is elvégezhető lenne, ahhoz egyrészt nem áll elég adat rendelkezésre, másrészt makroszinten az egyes vállalatok gazdálkodásbéli különbségeit nem lehetséges figyelembe venni. Magyarországon a megfelelő pontosságú mérést az ÁFSZ „Bértarifa” felvétele teszi lehetővé. Ebből megmérve a bérugalmasságot, azt találjuk, hogy az határozottan negatív, mínusz 50 százalék körüli. Ha kihasználjuk azt, hogy a „Bértarifa” felvétel nemcsak vállalati, hanem munkavállalói szinten is szolgáltat mikroadatokat, megmutathatjuk, hogy ez az átlagos rugalmasság a képzett és képzetlen munkavállalókra vonatkozó egyéni rugalmasságok kombinációja.

Már az 50 százalékos negatív rugalmasság is nem elhanyagolható veszteségeket jelez a foglalkoztatás csökkenése által. Ez az eredmény azonban tovább fokozódik, ha figyelembe vesszük, hogy a képzetlen munkavállalók esetében az a rugalmasság még nagyobb, a vállalkozások erőteljesebben reagálnak a béremelésre. Tovább árnyalja a képet, hogy tudjuk: a kisvállalatok érzékenyebbek a bérváltozásokra, jellemzően több képzetlen munkavállalót foglalkoztatnak, és a minimálbérhez közeli

bérért jellemzően az elmaradott térségben foglalkoztatnak embereket. Ha csak a bérek változását tekintjük, a negatív hatás nem látható, mert a fejlettebb térségekben található és nagyobb vállalatok esetében tapasztalható béremelkedés (a várt hatás) elfedi azt. A részletes eredmények arra utalnak, hogy a várakozásokkal szemben a minimálbér emelésének volt negatív hatása, ami alapvetően nem a bérek, hanem a foglalkoztatás terén, és nem annak szintje, hanem az abban megjelenő egyenlőtlenségek növekedésében csapódott le.

7.6. A mikroadatok hozzáféréseinek költség-haszon elemzése, különös tekintettel az adatvédelem költségeire

A mikroadatok hasznait az előbbi részekben alaposan körüljártuk: rámutattunk, hogy a mikroadatok kutatói használata olyan kérdések megválaszolását teszi lehetővé, amelyek rendkívüli gazdaság- és szociálpolitikai jelentőséggel bírnak. Ideje, hogy részletesen megvizsgáljuk ennek az árát, elérhetőségük költségeit is. Ezek a költségek szorosan kapcsolódnak a mikroadatok felhasználásának fázisaihoz: (- Adatfelvétel)

- Adatfelvétel részletes metaadatainak létrehozása
- Adatjavítás, tisztítás
- Adatjavítás, tisztítás metaadatainak létrehozása
- Felhasználói adatbázis leválogatása
- Anonimizálás
- Összekapcsolás

Noha az adatok felvétele elsüllyedt költség (a hozzáférés engedélyezésétől függetlenül kifizette az adatgazda), a felvétellel kapcsolatos metaadatok létrehozásáról már korántsem mondhatjuk nagy biztonsággal ugyanezt. Könnyen lehet, hogy számos, az adatok értelmezéséhez szükséges információ nem létezik papíron, azokat az intézményi folklór, vagy szélsőséges esetben egyetlen ember memóriája őrzi. A metaadatok iránti igényt tehát sok esetben éppen a nyilvánosságra hozatal teremti meg. Adott esetben elképzelhető az is, hogy ilyen dokumentumot az adat létrehozásában illetékessel együttműködésben harmadik személy, például az adatbázist felhasználó kutató, hozzon létre.

Az adatok javításának és tisztítását csak az dokumentálhatja, aki végzi is a feladatot. Szűkös források mellett ez a feladat komoly terhet jelenthet, ugyanakkor minőségbiztosítási céllal külső felhasználás nélkül is prioritást élvezhet

A modern adatbázisok kivétel nélkül úgynevezett relációs adatbázisban foglalnak helyet. Ezek felépítése a megvalósítástól függően eltérő lehet, lehetséges az is, hogy egy logikai egységbe összeállított adatbázis fizikailag különálló adattáblákban foglal helyet. A fizikai szerkezettől eltekintve abban viszont mind megegyeznek, hogy kezelésük standardizált, szakértő számára a mintavétel, vagy bizonyos részhalmazok pusztá leválogatása, ha számításokat nem kell végezni, nem jelent komoly feladatot.

A magyar Avtv. és a legtöbb nyugati ország vonatkozó jogszabálya megköveteli, hogy az adatok használata a személyiségi jogokat tartsa tiszteletben. Ez kétféleképpen történhet meg. Az első megoldás szigorúan jogi természetű: az adatokhoz annak kezelője/elemezője teljes hozzáférést kap, mindent „lát”, de nyilatkozat aláírásával kötelezi magát arra, hogy az adatokkal nem él vissza. Ilyen státuszúak például a statisztikai hivatalok munkatársai, és több országban ilyen státuszt kaphatnak kutatók is bizonyos feltételekkel. A második, jogilag kisebb felelősségvállalással járó megoldás az adatok anonimizálása. Anonimizáláson itt mindazokat az eljárásokat értjük, amelyek megszüntetik a kapcsolatot az tárolt adatok és a között a személy között, akire vonatkoznak. Anonimizálásnak tekintjük például azt is, ha az adatgazda véletlen zajt ad hozzá az adatokhoz.

Az anonimizálás első lépése az embereket egyértelműen azonosító jelzés elhagyása az adatbázisból. A statisztikai céllal felvett adatbázisok esetében ezeket az adatfelvétel után automatikusan törölni kell, ha pedig a vizsgálat az egyének időbeni követésére épül, az azonosítókat az adatoktól elkülönítve, szigorúan elzárva tartják. Nem igaz ugyanez az adminisztratív adatforrásokra: ott mindenképpen szerepel az egyének valamilyen, az államigazgatásban használatos azonosítója, amilyen a TAJ

szám, vagy az adószám. Bár ezeknek az eltávolítása többletfeladatot jelent, triviális műveletről van szó, így költségétől eltekinthetünk.

Az anonimizálás második lépése az adatbázis változóinak átvizsgálása abból a szempontból, hogy egyes kombinációi nyújtanak-e lehetőséget egyének azonosítására, azaz személyes adatnak tekinthetők-e. Az egész országra kiterjedő adatbázisból származtatott mintában az ilyen jellegű beazonosítás veszélye meglehetősen csekély, főleg, ha a területi azonosíthatóságot korlátozzák (mondjuk kistérségi szintre). Ha ez sem lenne elég, az adatok a legtöbb esetben összevonhatók és zavarhatók olyan formában, hogy az a kutató munkáját ne tegye lehetetlenné, az rosszhiszemű felhasználóét viszont igen.

Igazán sok ismerv vagy/és érzékeny adat esetében az anonimizálást végző adatkezelő döntési helyzet elé kerül: mit tegyen, ha nem lehet 100 százalékgig biztos abban, hogy a válaszadó nem lesz azonosítható. Egyik lehetősége természetesen az, hogy az adatokat visszatartja, de a korábbiakban kifejtett okoknál fogva ez a lehetőség mindenképpen szuboptimális eredményhez vezet. Léteznek ugyanakkor olyan adattranzformálási lehetőségek, amelyek által jelentősen csökkenthetők az adatokkal való visszaélésekből származó várható veszteségek, miközben az adatok kutatási felhasználhatósága jelentősen nem változik. A következő mesterséges példa ezt próbálja egyszerűen illusztrálni.¹⁶²

Tegyük fel, hogy az adatgazda több ezer megfigyelt személyről rendelkezik adatokkal azok jövedelméről, pár demográfiai jellemzőről (például kor, nem, családi állapot), és arról, hogy mekkora összeget fordítanak havonta nyugdíjbiztosításra. Ez az adatbázis alkalmas arra, hogy egy kutató (jóhiszemű felhasználó) számszerűsítse a megtakarítási hajlandóság kialakulását befolyásoló tényezők hatásait, amely kutatási eredményt társadalmilag hasznosnak gondoljuk. Ugyanakkor attól félünk, hogy amennyiben rosszhiszemű felhasználó kezébe kerül ez az adatbázis, ő a demográfiai adatok alapján képes lenne azonosítani az egyes

¹⁶² A módszer teljes bemutatására és a hozzá kapcsolódó matematikai módszerekre lásd Duncan (1999), valamint Duncan és szerzőtársai (2001).

személyeket, és így megismerné az érzékeny adatnak minősülő jövedelmüket, amit károsnak gondolunk.

Egy triviális lehetőség ennek a kárnak az elkerülésére az, hogy a jövedelemadatokat az adatgazda nem adja ki, ezzel ugyanakkor gyakorlatilag megakadályozza, hogy a kutatási kérdést a jóhiszemű felhasználó megválaszolhassa, ugyanis a jövedelem igen jelentős tényező a megtakarítási hajlandóság alakulásában. Ez az adatvédelmi lehetőség nulla kárt okoz adatvédelmi oldalon, ugyanakkor jelentős elmaradt haszonhoz vezet a kutatói oldalon.

Vizsgáljuk meg csak verbálisan annak az adatvédelmi intézkedésnek a hatását, amikor az összes jövedelmi adatot egy nulla várható értékű, de igen nagy szórású véletlen zaj hozzáadásával torzítjuk, és ezt az adatbázist adjuk közre. A rosszhiszemű felhasználó amennyiben be is tud azonosítani egy egyént, mindössze a rá vonatkozó egyetlen torzított adat ismeretében tud becslést adni annak jövedelmére, ami a nagy szórású zaj miatt igen pontatlan lesz. A jóhiszemű kutató ezzel szemben az adatok sokaságát vizsgálja, és minél nagyobb az adatbázis, annál kisebb lesz a véletlen zaj torzító hatása a megfigyelt általános jelenség becslésekor. A zaj hozzáadásával így a visszaélések által történő károkozás kockázata jelentősen, míg a felhasználhatóság csak kis mértékben csökken. Matematikai módszerekkel az adatbázis nagyságának, valamint a kutatási eredményektől várható haszon, illetve a visszaélés okozta várható kár ismeretében meghatározható a hozzáadott zaj optimális nagysága, ennek ismertetésétől itt most eltekintünk.

Noha a fenti példa igen speciális esetet vesz alapul, rámutat, hogy az adatok információ elvesztésével járó anonimizálása nemcsak az anonimizálással járó költségterhet vonja maga után, de a realizálható hasznokat is csökkenti. Minél több információt veszít az adatokat használó kutató, annál kevésbé lesz az értékes számára, és a munkájából fakadó externális hatás révén az egész társadalom számára az adat. Ha feltétlenül szükség van az információ csökkentésére, az a

használt statisztikai eszköz ismeretében megtehető úgy, hogy a kutatási célú felhasználást ne érintse hátrányosan, a rossz szándékú felhasználás hasznát viszont erősen csökkentse. Ha erre valamilyen oknál fogva nincs mód, vagy az adatok minőségében a kutatási célú felhasználás számára elfogadhatatlan romlás áll be, a megoldást nem az ilyen típusú adatvédelem, hanem például a hozzáférés erőteljes szűrése jelentheti.

Az azonosítóknak azonban nem csak az eltávolítás lehet a sorsa, hanem egymástól független adatbázisok összekapcsolásának eszközévé is válhat. Az adminisztratív adatok összekapcsolásának azonban Magyarországon komoly intézményi akadályai vannak. 1991-ben az Alkotmánybíróság alkotmányellenesnek nyilvánította a személyi szám használatát¹⁶³, és elrendelte, hogy a különféle nyilvántartásokban más és más azonosítót alkalmazzanak. Ennek célja az volt, hogy a különféle adatbázisok egymással ne legyenek összekapcsolhatók az államigazgatási munka során. Annak ellenére, hogy az intézkedés nem a hatékonysági szempontokat helyezte előtérbe, a két fontos azonosító, a TAJ- és az adószám közötti átjárás elméleti lehetősége megmaradt. A Belügyminisztérium Központi Nyilvántartó és Választási Hivatalának (BMKÖANYVH) birtokában van az a „kulcs”, amellyel a kapcsolat megteremthető.

Ha valaki olyan kutatást szeretne végezni, amelyhez szükséges a vizsgált népesség egyéni személyi jövedelemadó adatait gyógyszerfogyasztási adataival összekapcsolni (természetesen továbbra is szem előtt tartva az adatvédelmi szempontokat, az egyének azonosítását lehetővé tevő adatot elhagyva), azt az adatvédelmi törvénnyel összhangban megteheti. Ehhez azonban a BMKÖANYVH közreműködésére van szükség, aki megbízható harmadik félként mindkét adatgazda számára elküld listát a véletlenszerűen kiválasztott személyek TAJ illetve adószámaival (természetesen mindkét esetben csak az egyiket), illetve az azokhoz kapcsolt, újonnan generált azonosítóval együtt. Ezáltal a BMKÖANYVH úgy teremti meg a kapcsolatot a két azonosító között, hogy rajta kívül (akinek ezt a jogszabály megengedi) azok sehol sem voltak összekapcsolva. A két intézmény mintát vesz,

¹⁶³ 15/1991. (IV. 13.) AB határozat, ABH 1991, 40, 42.

szükség szerint anonimizálja azt, elhagyva és/vagy összevonva az azonosítást segítő változókat. A keletkezett két adatbázis már közérdekű adatnak tekinthető, hiszen állami intézmény kezelésében van, és nem személyes adat. A felhasználó így már képes összeilleszteni a két adatbázist és felhasználni a kívánt célra.

A hazai tapasztalatokat bemutató fejezetben példát is fogunk hozni ilyen jellegű adatkérésre. Ismét a hazai gyakorlat sajátossága az, hogy az adatkérés nem valósulhatott meg, a kért feladatot a BMKÖANYVH nem teljesítheti, jogszabályi megokoltság hiányára hivatkozva.

7.6.1. Egy közvetlen haszon: a kutatói visszajelzés

A kutatói adatfelhasználás haszna nemcsak a kutatók által létrehozott származtatott adat, ismeret közzétételével realizálódhat. Az adatfelhasználás, és különösen a mikroadatok kifinomult felhasználása rendkívül sok technikai jellegű teendővel jár. Az adatok igen zajosak lehetnek, szerkezetük nem feltétlenül segíti a mikroszintű elemzést, elképzelhető, hogy dokumentációjuk nem teljes, és még sorolhatnánk. Mint azt már említettük, a mikroadatok sokszor olyan környezetben készülnek, ahol ezen problémák megoldására elégtelen a késztetés.

A főként táblázatok készítésére specializálódott tevékenység nem feltétlenül követeli meg az adatok olyan pontosságát, mint egy igen összetett és finom elemzés. Ez persze nem jelenti azt, hogy ha valamilyen oknál fogva már eleve rendelkezésre állna a létezőnél jobb, alkalmasabb, tisztább adatbázis, azt nem használnák azok, akik főként táblázatokkal dolgoznak. Az ehhez a szükséges munkákat nem azért nem végzik el, mert nem tudnak a problémák létezéséről, vagy mert nem rendelkeznek a megfelelő tudással. Az ok egyszerűen az, hogy a szükséges munkabefektetés számukra nem térül meg.

Ha a mikroadatok hozzáférhetőek kompetens kutatói kör számára, az adatgazda értékes munkaerőre tehet szert. Megfelelő, akár szóbeli információk alapján az

igényes kutatók a legtöbb számukra fontos hibát fel fogják térképezni és legjobb tudásuk szerint ki is javítják. Minél alaposabb elemzést készít egy kutató, annál alaposabb javítást remélhet tőle az adatgazda. Ha sikerül jó kapcsolatot kiépíteni a felek között, az mindkettőjüknek hasznára válik, hiszen nemcsak a kutató, de az adatgazda is igényesebb munkát végezhet. Ez nem csak az aktuális munkát segítheti, de későbbi adatkérések esetében jobb minőségű adat átadását teszi lehetővé, ami végeredményben az adatgazda szakmai reputációját növeli. Az ide vezető úton természetesen az első lépés a hozzáférhetőség biztosítása.

7.6.2. Egy rejtett költség: intézményi szavahihetőség és stabilitás

Az adatok nyilvánosságának eddig kizárólag egy csatornán, a tudás gyarapodásán keresztül érvényesülő hatásáról beszéltünk. Amellett érveltünk, hogy ez a hatás kizárólag pozitív lehet. Ezt fenntartva, most egy másik, kontextuális szempontot vezetünk be, amely hatását a gazdaság szempontjából alapvető fontosságú intézmények működését befolyásolva fejti ki.

A mikroadatok nyilvánossága nemcsak korábban nem elérhető tudás reprodukálására alkalmas, de olyan eredmények újragondolására is, amelyek már léteznek. Vegyük példának a fogyasztói árindexet! A fogyasztói árindex az a szám, amely az „infláció” fogalmát a valóságban megjeleníti. Széles körben használják, a reálbérek számításától kezdve a különféle projektek értékelésén keresztül különböző állományok, például nyugdíjalapok tőkeállományának meghatározására. A fogyasztói árindex tehát igen széles körben használt mérőszám, gyakorlatilag minden gazdasági szereplő használja, és mindegyik tudja a másiktól, hogy ezt használja.

A magyar fogyasztói árindexet a KSH meghatározott módszertan szerint készíti, az átlagos fogyasztó kiadási szerkezetét megjelenítő 150 termék árváltozásainak átlagaként, amely átlagban a súlyokat a lakosság (tárgyidőszak előtti 2. év) kiadási szerkezetében mért arányok szolgáltatják. Az árváltozásokat boltokban tett

megfigyelések alapján gyűjtik össze, a kiadási arányokat pedig lakossági adatfelvételtől számítják. A jelenlegi módszer rendkívül átgondolt, nemzetközi összevetésben is igen színvonalas; kialakítását komoly szakmai munka előzte meg. Ezzel együtt ez is, mint bármilyen módszer, számos olyan döntést tartalmaz, amit ha valaki feltétlenül ragaszkodik hozzá, önkényesnek, „kijavíthatónak” tarthat.

Ilyen esetleges döntés eredménye például az egyes rész-termékek (az úgynevezett „reprezentánsok”) aránya a 150 termékben. Mivel a háztartási adatfelvételen a végrehajthatóság érdekében nem kérdeznek rá a kiadásokra teljes részletességben, a kiadási arányok olyan absztrakt fogalmakra vonatkoznak, mint a „sertéshús”, nem a konkrét reprezentánsokra. Ezt az információhiányt áthidalandó, a statisztikusok a konkrét esetben azzal a feltevessel élnek, hogy a kiadásokban olyan arányban vesznek részt az egyes húsok, amilyen súlyarányban azok egy tipikus disznó testében megtalálhatók.

A fogyasztói árindex nem mindenhol olyan jó minőségű, mint Magyarországon. Az Egyesült Államokban például rendkívüli port kavart a Boskin bizottság jelentése, amely rámutatott, hogy az USA fogyasztói árindexében igen jelentős, 1 százalékpontos torzítás lehet annak következtében, hogy a súlyokat csak igen hosszú idő után revideálják. A korrekció jelentős mértékben befolyásol több, igen lényeges mutatót, például az államadósság és a nyugdíjkötelezettség nagyságát. Az ilyen vizsgálatoknak tehát igen nagy jelentőségük van: külső szakértő kontrollt biztosítanak olyan intézmények felett, amelyek munkája saját logikájukat követi, és amelyek hatóság jellegű működését „nem szokás” kritizálni. Az adatok nyilvánossága bármikor lehetőséget biztosít ilyen kritikai megfogalmazására.

Ha azonban az elemi ár-adatok mindenki számára hozzáférhetőek, egy vállalkozó kedvű piaci szereplő bármikor hozzáláthat, hogy alternatív fogyasztói árindexet kreáljon, „felülvizsgálja” a „hivatalos” adatot. Függetlenül attól, hogy mennyire gondoljuk valószínűnek ezt a lépést, érdemes eljátszani a gondolattal, mi történik egy ilyen esetben. A piac szereplői elbizonytalanodnak, egy része elkezd az új indexet használni. Bár mindenki pontosan ki tudja számolni, hogy például egy adott

befektetés megéri-e az érte kért összeget, az egyik indexet használó csoport ugyanabban a helyzetben kicsit másként fog dönteni, mint a másik. A fogyasztói árindex természetesen csak egy a sok lehetőség közül: felülvizsgálható a bruttó hazai terméktől elkezdve az államadósságig sok minden.

Az, ha fontos mérőszámok megbízhatósága megkérdőjeleződik, nehezen belátható következményekkel jár. Az, hogy nehezen belátható, alapvető fontosságú: a következmény nem feltétlenül tragikus, mint ahogy a kapott mérőszám sem feltétlenül sokban tér ez a „hivatalos” változattól. Az azonban bizonyos, hogy ha egyszer meginog a bizalom egy ilyen, alapvetőnek számító statisztikában, azt igen nehéz visszaszerezni. Mindebből az építő kritika és az adatok teljes nyilvánosságának kívánatosságával ellentétes következtetésre juthatunk. A lehetséges megoldásokra a nemzetközi gyakorlat áttekintése után térünk vissza.

8. A mikroadatokhoz való hozzáférés nemzetközi tapasztalatai

A következő fejezetben áttekintjük a mikroadatok szolgáltatásának nemzetközi gyakorlatát, néhány tipikus intézményi modell bemutatásával. Ez mintegy aláfestésül szolgál a korábban részletezett, az adatszolgáltatás kapcsán felmerülő lehetőségekhez, bemutatja azok gyakorlati megvalósulását. Mint látni fogjuk, ezek igen különbözőek lehetnek, és bizonyos szempontból egy adott ország társadalmi, gazdasági, ha úgy tetszik piaci szerkezete is behatárolja, hogyan valósulhatnak meg az elvek. Ezért a legértékesebb következtetéseket nem úgy vonhatjuk le a modellekből, ha megpróbáljuk egyiket vagy másikat reprodukálni, teljes egészében magunk elé állítani, hanem meghatározó tulajdonságaik mentén haladunk, a megvalósult gyakorlatok különböző elemeit összehasonlítva. Az előző részekben tárgyalt elméleti következtetéseket a következő intézményi modellekkel illusztráljuk: az északi országok (egymáshoz igen hasonló) módszereit, majd az Amerikai Egyesült Államok és Nagy-Britannia példáját vesszük sorra a nemzeti szintű adatszolgáltatási gyakorlat kapcsán, végül kitérünk két szupranacionális intézmény, az Európai Unió és az IMF ezirányú szabályalkotásaira, irányelveire is.

Természetesen a világ számos országában áll rendelkezésre fejlett adatgyűjtő, adatszolgáltató intézmény, és abban is biztosak lehetünk, hogy rengetegféle gyakorlat létezik arra vonatkozóan, hogyan juttassák, vagy ne juttassák el az adatokat a felhasználókhoz. Mi mégis ezt a hármat (illetve a két nemzetközi intézmény standardját) emeljük ki, több okból is. Egyrészt mert ezek mentén egymástól lényegesen eltérő, mondhatni tipikus szerkezetekhez jutunk. Látunk példát egy kis méretű, erősen centralizált piacra, egy igen nagy és jelentősen decentralizált jellegűre, valamint természetesen létezik a kettő közötti átmenet is. Választásunk másik oka, hogy mindhárom esetben letisztult, már (a mikroadatok kezelésének terén is) hagyományokkal rendelkező adatfeldolgozó rendszerről van szó, s valamilyen módon mindenütt lehetővé teszik külső (nem a statisztikai hivatal alkalmazásában álló) szereplők, tehát akár kutatók, akár a piac hozzáférését is. Harmadik okként pedig az sem elhanyagolható, hogy a mikroadatok kutatásával

foglalkozó közgazdászok, és más kutatók számára (elmondásuk alapján) általában ezen országok adatbankjai bizonyulnak a legfontosabb, legkeresettebb forrásnak.

A nemzeti szintű modellek után (és legtöbb esetben egyébként valóban országhatárok között maradó szolgáltatásról van szó) pedig természetesen nem hagyható ki a vizsgálat köréből a nemzetközi standardok áttekintése sem. Ebből is sok született már, itt azonban csak néhányra térünk ki: az IMF és az Európai Unió vonatkozó ajánlásaira, szabályozására melyek közül, az Unióhoz való csatlakozásunk óta különös aktualitása miatt is, leglényegesebb az Európai Bizottság bizalmas adatok szolgáltatására vonatkozó irányelve. Figyelembe veendő az EU Adatvédelmi irányelve is, amelynek célja az adatvédelem megerősítése az adatok tagállamok közötti szabad áramlása lehetőségének biztosítása mellett. Ezt nem tekintjük külön át, mert vizsgálódásunk tárgya nem az egyén, a magánélet és a személyes érdekek védelme, hanem az adatok felhasználhatósága. Az adatvédelmet egyébként valamennyi általunk vizsgált adatfeldolgozó intézmény kiemelkedően fontosnak tartja, ennek érdekében különböző korlátozásokat alkalmaznak, de általában igyekeznek megtalálni az egyensúlyt az egyéni integritás fenntartása és az adatok felhasználhatósága között.

Először megvizsgáljuk, hogy milyen intézményi keretek között mozognak az adatszolgáltató szervezetek. Kitérünk a jogi környezet bizonyos elemeire is, hiszen az egyes országok, vagy szervezetek rendelkezései stabil keretbe helyezik ezeket a szolgáltatásokat, az egyik oldalról biztonságot a másiktól korlátokat teremtve. Ehhez kapcsolódik részben az adatvédelem kérdése is, amit kétségtől alapjaiban meghatároznak a vonatkozó jogszabályok, ugyanakkor gyakorlati problémákat is felvet.

Ezután példákat hozunk arra, hogyan teszik ezekben az országokban hozzáférhetővé a mikroadatokat: milyen módon, milyen felhasználó körben és milyen céllal. Ennek kapcsán láthatjuk majd az korábban áttekintett árazási lehetőségek alkalmazását ezekben a modellekben. Ennél a pontnál kitérünk majd az egyéb, nem árjellegű költségekre is, a felhasználókkal szemben támasztott

elvárásokra, akadályokra, és a különböző diszkriminációs stratégiákra is. Külön részt szentelünk a társadalmi környezetnek is, mivel az intézmények és a kutatók közötti együttműködés változataira is számos példa akad. Ezáltal tágabb áttekintést nyerhetünk az adatgyűjtéstől a feldolgozáson át a nyilvánosság számára is elérhetővé tételig húzódó folyamat jellemző változatairól, az ebben rejlő fejlesztési lehetőségekről, szemlélet- vagy éppen hatékonyságbeli különbségekről.

Az egyes modellek összevetésekor megnyilvánuló különbségek részben olyan komplex okokra vezethetők vissza, mint a történeti-társadalmi környezet, intézményi hagyományok, az adatok iránt megnyilvánuló külső, tehát intézményen kívüli kereslet, a potenciális felhasználók érdekérvényesítő képessége, láthatósága, tudományos életbeli ereje (olyan értelemben is, hogy mennyire tudják az egész társadalom hasznára fordítani munkájukat). Minderre azonban csak háttérelemként utalunk, célunk elsősorban a megnyilvánuló, mindennapi gyakorlat közvetlen okainak, összetevőinek feltárása. Ezt a gyakorlatot úgy tekintjük, mint példát arra, hogyan lehet az adatok iránt megnyilvánuló (tudományos célú) kereslet, a létrehozásukkor felmerülő költség, illetve (mintegy további költségként) a természetükben, tehát például bizalmas jellegükben rejlő jogi/erkölcsi természetű korlátok rendszere által meghatározott térben sikerrel mozogni.

8.1. Intézményi környezet, a mikroadat-szolgáltató piacok szerkezete

A felsorolt modellek, mint azt a bevezetőben is láhattuk, az intézményrendszer struktúrájában jelentősen eltérnek egymástól, három különböző modellt testesítenek meg. Az *első* az északi országokban jellemző, centralizált nyilvántartással és statisztikai adatkezeléssel rendelkező kis méretű ország példája. Ebben a csoportban majd mindegyik országról elmondható, hogy egyetlen központi statisztikai hivatal van, ami egyben a kiterjedt nyilvántartásokat (népesség, foglalkoztatás, adózás, álláskeresés, nyugdíj, üzleti szféra) is vezeti. Igen szoros tehát az intézményi összefonódás, ami a regiszterek azon tulajdonságával együtt, hogy egymáshoz nagyon jól illeszkednek, különlegesen jó rálátást biztosíthat a társadalom és a gazdaság egészére. Ennek a leginkább az északi országokra

jellemző tulajdonságnak az okát a jóléti állami struktúrájában, a kiterjedt szociális és ellátási rendszer működtetésének „információéhségében” kereshetjük, de az is elmondható, hogy ezek az országok évszázados hagyományokkal rendelkeznek a statisztikai rendszer fenntartása terén.

A *második* intézményi modell, amelynek példája az Egyesült Államok, egy különösen sokelemű, decentralizált rendszer, ahol a piacon nincsen kiemelt központi szereplő, hanem az egyes kutatások végrehajtói vagy az adatok feldolgozói függetlenek, sok esetben mind a finanszírozást, mind az irányítást vagy a döntéshozást tekintve. Jellemző, hogy külön intézmény végzi a legjelentősebb (és széles körben felhasznált) felvételnek, a tízévenkénti népszámlálást és az ország legfontosabb gazdasági mutatóinak előállítását, közzétételét. Ennek a modellnek az is sajátossága (mind az elsőhöz, mind a harmadikhoz képest), hogy igen nagy sokaságot vizsgál, hiszen az Egyesült Államok népessége jóval meghaladja a többi országét, ugyanakkor az alkalmazott mintaelemszám és a módszerek hasonlóak (például az angol munkaerő-felvétel mintájában 60 000, míg az amerikaiakban, ahol ebből több is van, 30-50 000 háztartás szerepel), így a felvételek reprezentativitási tulajdonságai bizonyos kérdésekben különböznek. Alighanem ez is oka annak, hogy a felvételek párhuzamosan futnak, a földrajzi, demográfiai és tematikus lefedettségű területek elhatárolása nem olyan egyértelmű (a foglalkoztatottság és a jövedelem felmérése például számos kutatásban megjelenik, más-más súllyal és reprezentativitással).

Természetesen minden statisztikai rendszer rendelkezik azzal a különleges tulajdonsággal, hogy az egyes intézmények fokozott együttműködése nélkül sokkal kisebb hatékonyság érhető el. Praktikusan ennek két fontos következménye van, ami behatárolja az intézményi önállóságot. Az egyik az, hogy a független intézményeknek is szükségük van bizonyos más adatokra, például a különböző nyilvántartásokból, vagy a népszámlálásból, ezért ezek mindenképpen központi jelentőségűek. A másik következmény, részben éppen az előző miatt, de részben az adatok különleges értéke okán, hogy még egy ilyen decentralizált modellben is meglehetősen erős az állami jelenlét a szabályozásban és a finanszírozásban is. A

különbség az egyes modellek között inkább egyrészt abban látható, hogy ez a központi „adatellátó” mennyire erős szervezet, azaz mennyire jelentősek az egyéb kutatásaik, mennyire van hatással (akár szabályozáson, bürokratikus viszonyrendszeren keresztül, akár ármeghatározás által) a többi szereplőre. A második modellnél ez a hatás kicsi. Másrészt pedig az állami beágyazottság is különböző mértékű lehet, itt gyakran csak áttételes a kapcsolat, és leginkább a finanszírozásra korlátozódik.

A harmadik típus szerkezetét tekintve az első kettő között helyezkedik el. Ennek jellemző példájában, Angliában, létezik ugyan Nemzeti Statisztikai Hivatal, ez azonban már elsősorban kormányzati adatgyűjtési funkciót lát el (bár emellett ide tartozik néhány jelentősebb társadalmi felvétel). Ez egyben a népesség-nyilvántartó szervezet is és szoros kapcsolatban áll az adóhivatallal. Ezen kívül megjelennek más magán és állami alapítású szervezetek, autonóm intézmények, amelyek szintén végeznek kutatásokat, jóllehet általában ezek is jelentős állami támogatással működnek, azért meglehetősen nagy önállósággal és pénzügyi függetlenséggel rendelkeznek. Ebben az esetben jóval kevesebb a szereplő, mint az amerikai típusú modell esetében, ugyanakkor az is igaz, hogy az intézményi decentralizáció meglehetősen nagy mértékű összefonódással párosul, ami mégis az egységesség felé mutat. Ezt főleg ott láthatjuk, ahol ennek méretgazdaságossági vagy államigazgatási előnyei vannak: egyrészt az adattárolás terén figyelhető meg, hogy egyetlen központi archívum kezelésében áll a legtöbb jelentős felvétel adatállománya, másrészt a finanszírozás is általában (részben) ugyanattól a kormányzati szervtől ered.

Az összefonódásnak több oldala van. Mondhatjuk egyrészt, hogy ezt a magas állami részvételi arány okozza, vagyis ez is az irányítás egy módja, de úgy is megközelíthetjük, hogy erősebb a kooperáció, éppen a már említett méretgazdaságossági előnyök kihasználása érdekében. Ez bizonyos fajta rugalmasságként is értelmezhető, hiszen ezek az intézmények nem csak saját kereteiken belül tevékenykednek, hanem elképzelhetőek projekt-szerű, időleges együttműködések is, például bizonyos kutatási módszerek, területek összehangolt

fejlesztésére, a másodlagos adatfeldolgozás, felhasználás elősegítésére. Ennek eredményeképpen a különböző adatfelvételek, kutatási területek sokszor különböző csatornákon keresztül, több irányból is elérhetőek.

8.1.1. Az északi országok statisztikai rendszere

Az északi országok közül, mint említettük, mindegyik rendelkezik központi vagy nemzeti statisztikai hivatallal, ami Dániát kivéve egyben a népesség-nyilvántartó is. Svédország esetében van egy kis eltérés, néhány éve 24 más hatóság, intézmény is tárol adatokat (például kikerült így az oktatási, és az egészségügyi statisztikák kezelése), de az adat-előállításban itt is a központi hivatalnak van a legnagyobb szerepe. Az ezekben a hivatalokban tárolt adatok tehát szorosan összefüggnek a különböző regiszterekkel (népesség, foglalkoztatás, adózás, álláskeresés, nyugdíj, üzleti szféra), különösen hatékony rendszert alkotva, minthogy jellemzően az összes regiszterben ugyanazok az azonosítók tartoznak az egyes személyekhez, vállalkozásokhoz, intézményekhez és lakcímekhez.

8.1.2. Az Amerikai Egyesült Államok statisztikai rendszere

Ha központi szerepű intézményt keresünk, akkor kettőt említhetünk, melyek legalábbis abból a szempontból kiemelten kezeltnek számítanak, hogy mindkettő valamely kormányzati szerv felügyelete alá tartozik. Az egyik a népszámlálást készítő *Census Bureau (CB)*, ami számos más, széles körben felhasználható lakossági és gazdasági adatfelvételt is futtat. Az egyik legjelentősebb ezek közül, 50.000 háztartásból álló mintája folytán, a *Current Population Survey (CPS)* elnevezésű munkaerő-felvétel, aminek feldolgozásában egyébként a *Bureau of Labor Statistics* (lásd később) is részt vesz. De a *CB* készíti a fogyasztási költségvetési felvételt (*Consumer Expenditure Survey*) illetve egy jövedelemről és a kormányzati programokban való részvételről szóló panelfelvételt is (*Survey of Income and Program Participation - SIPP*). A másik központi szereplő a gazdasági mutatók előállítását végző *Bureau of Labor Statistics*, ami szintén foglalkozik

longitudinális lakossági felvételekkel is (*National Longitudinal Surveys*, kiemelkedik ezek közül az egyébként szintén munkaerő-piaci fókuszú *National Longitudinal Survey of Youth*, azaz az ifjúsági longitudinális felvétel). Ezen kívül tevékenységével kitűnik még a *National Bureau of Economic Research (NBER)*, ami nagy múltú közgazdasági kutatóintézet, ám nem rendelkezik direkt adatgyűjtő funkcióval, főleg kvantitatív modellek, valamint a kormányzat és a gazdaság viszonyának vizsgálata kapcsán használnak és tisztítanak empirikus adatokat.

A nagy, egyértelműen állami háttérű intézmények mellett vannak egy-egy nagyobb kutatásuk miatt említést érdemlő kisebb kutatóközpontok is. Ezek közül egyértelműen kiemelkedik a Michigani Egyetem társadalomtudományi kutatóintézete (*Institute for Social Research*): itt készül a *Panel Study of Income Dynamics (PSID)* nevű egyedülállóan kiterjedt, 36 éve (majdnem) ugyanazzal a mintával futó jövedelmi panelfelvétel, valamint az időskorúak egészségi állapotát, szociális és vagyoni helyzetét vizsgáló *Health and Retirement Study (HRS)*.

8.1.3. Az Egyesült Királyság statisztikai rendszere

Az Egyesült Királyság statisztikai rendszerében erős az állami jelenlét. A Nemzeti Statisztikai Hivatal (*Office of National Statistics - ONS*) egyik fő feladata a kormányzati statisztikai szolgálat működésének összefogása, és bizonyos szinten a statisztikai tevékenység minőségi ellenőrzése. Ez a hivatal hozza nyilvánosságra az ország főbb gazdasági és társadalmi mutatóit, ez készíti a tízévenkénti népszámlálást és ennek része Anglia és Wales népesség-nyilvántartása (*General Register Office for England and Wales*) is. Mintavételes adatfelvételeik szintén alapvetően kormányzati célokat szolgálnak: a jelentősebbek közül ide tartozik a munkaerő-felvétel (*Labour Force Survey*) bizonyos fogyasztási és táplálkozási felmérések (*Expenditure and Food Survey*, *General Household Survey*) illetve például 14-féle kisebb-nagyobb egészségügyi állapottal és ellátással kapcsolatos felvétel is. Gazdasági területen kevésbé tevékenykedik a hivatal, mindössze a

vállalati szektor éves, kötelező válaszadáson alapuló mintás lekérdezését (Annual Business Survey) végzik itt.

A piacon jelenlévő más magán és állami alapítású szervezetek, autonóm intézmények is jelentős állami támogatással működnek. A legfőbb finanszírozó szerv az Economic and Social Research Council (ESRC), ami mintegy 100 millió fontra rúgó (kormányzati eredetű) költségvetéséből támogatja a társadalomtudományok kutatóit, egyéni és intézményi szinten egyaránt. Ilyen támogatott intézmény néhány egyetemi kutatóközpont, például a kutatási téren igen jelentős tevékenységet kifejtő Essexi Egyetem longitudinális felvételeket készítő intézete, az Institute of Social and Economic Research (ISER).

Az ESRC később autonóm módon működő intézetek, kutatóközpontok alapításában is részt vesz. Kiemelkedő jelentőségű ezek közül a UK Data Archive (UKDA) nevű adatarchívum: tevékenysége az adatfeldolgozás folyamának minden állomására kiterjed: adatgyűjtéssel, -megőrzéssel és -szolgáltatással foglalkozik, hozzáférést nyújt több, mint 5000 adatállományhoz, köztük számos nagy adatfelvételhez, kutatási és oktatási célból. Ezek tematikus egységekre bontottak: vannak kormányzati adatfelvételek, longitudinálisak és egyebek, amiket szintén itt tárolnak, köztük nem utolsó sorban, a népszámlálás adatai. De elérési lehetőséget ad más, nem itt tárolt adatokhoz is, mint például egyes nemzetközi szervezetek (Nemzetközi Energia Ügynökség, OECD...) által készített kutatásokhoz.

Hasonlóan okból jött létre a Manchesteri Egyetem információs szolgáltató központja (Manchester Information & Associated Services - MIMAS), amely főleg oktatási és kutatási célú adatbázisokat kezel: folyóirat- és könyvtári állományok, tananyagok, linkek gyűjteménye és egyéb szolgáltatások (például tréningek, szakértői segítségnyújtás). Ennek egyik részlegén keresztül (Census Dissemination Unit) érhetők el a népszámlálás adatai.

Mindezen eddig említett intézmények longitudinális felvételek kezelésére, továbbfejlesztésére közösen létrehozták Economic and Social Data Service (ESDS)

nevű szolgáltatót, ami jó példa a már említett független kutatóintézetek közötti projekt-szerű együttműködésre. Célja hogy támogassa ezeknek a számos (akadémiai vagy akár piaci) forrásból származó adatoknak a tudományos életbeli és oktatásbeli felhasználását, elsősorban a társadalomtudományok területén. Ehhez kapcsolódóan a magánszférából is várják adatok beérkezését, vállalják ezek archiválását. Ez egy több tematikus egységből álló rendszer: itt található a főbb longitudinális felvételek (például a háztartáspanel), bizonyos kormányzati megbízásból készülő adatfelvételek (például a munkaerő-felvétel), kvalitatív adatbázis, de nemzetközi adatok eléréséhez is igyekeznek segítséget nyújtani.

Az intézmények közötti összefonódás tehát, a harmadik modell esetében erős, a különböző adatfelvételek, kutatási területek különböző csatornákon keresztül érhetők el, gyakran több irányból, ám alapvetően néhány intézményre szűkíthető az a kör, amivel a kutatók kapcsolatba kerülnek.

8.2. Jogi környezet

A statisztikai adatszolgáltatásról szóló jogalkotást több oldalról meg lehet közelíteni. Számunkra itt, a nemzetközi gyakorlatok bemutatása szempontjából ennek talán három aspektusa érdekes: a szabályozás tárgya, szintje és erőssége.

8.2.1.A szabályozás tárgya

Az első csoportosítás szerint fontos elkülönítenünk (többnyire egyébként is külön jelennek meg) az adatvédelemre és a statisztikai intézmények működésére vonatkozó szabályokat. Előbbiek az állampolgárok jogaival függenek össze, az egyéni érdek, integritás sértetlenségét helyezik előtérbe. Ilyen törvényt az általunk vizsgált modellekben mindenütt találunk, és az EU-nak is létezik ilyen irányelve. A másik típus a statisztikai intézményekre vonatkozó szabályozás, ilyenre is számos példa akad (a szabályozások minden szintjén), bár a fontossági sorrendben inkább

második helyet foglal el: van, ahol nem törvények, hanem irányelvek útján szabályozott ez a terület.

Elsőként tehát lássunk néhány példát az adatvédelmi intézkedésekre, amit voltaképpen máris két oldalról közelíthetünk meg, hiszen a bizalmas adatok védelme egyben a hozzáférés korlátozását is jelenti. Ez az általunk vizsgált országok jogalkotásában legszembetűnőbben az Egyesült Államokban jelenik meg, ahol három fontos törvény is vonatkozik egyazon területre, más-más megközelítés alapján adva keretet a kormányzati- és statisztikai szervek, intézmények működésének. Hasonló a helyzet Nagy-Britanniában is, ahol néhány éve ugyancsak készült egy törvény az információszabadságról, és ez 2005 januárjától hatályba is lépett, ám először (komplexebb felépítése miatt) az amerikai szabályozás modelljét vizsgáljuk meg kissé részletesebben.

Az Egyesült Államokban tehát a három fontos törvény a *Freedom of Information Act* (FOIA - az információ szabadságról), a *Privacy Act* (ami a szövetségi adatgyűjtő intézmények személyes adatokra vonatkozó gyakorlatáról szól) és a *Confidential Information Protection and Statistical Efficiency Act* (CIPSEA - azaz a bizalmas adatok védelméről és statisztikai hatékonyságról szóló törvény). Mindhárom az Egyesült Államok törvénykönyvének 5. fejezetében található. A továbbiakban ezeket, a szabályozás különlegessége és a szabad információáramlásnak szentelt figyelem okán kissé részletesebben is megvizsgáljuk.

Az első említett jogszabály (FOIA) célja, hogy biztosítsa minden egyén hozzáférését a szövetségi kormányhivatalok (így például a BLS, de nem a privát intézmények, sem pedig az egyes államok) adatállományaihoz (rekordjaihoz, ami itt tág értelemben használatos, mint adattárolási egység), bizonyos időhatárokon belül, illetve, ha ez nem lenne lehetséges, az intézmény megfelelő tájékoztatást adjon ennek okairól. Hangsúlyos a tájékoztatás, a hatálya alá tartozó intézmények hivatalos publikációk, tájékoztató anyagok készítésére is kötelezettek, ám nem minden adat esik ezen szolgáltatási kötelezettség hatálya alá. Természetesen kivételt képeznek a nemzetbiztonsági vagy kereskedelmi, gazdasági érdekből

titkosnak minősített adatok, az egyes szervezeteken belüli és azok közötti kommunikáció elemei, és a személyiségi jogokat érintő adatok. Mindazonáltal, ha egy rekord ilyen, a törvény hatálya alá nem tartozó adatot tartalmaz, ebben az esetben az adatot kezelő intézménynek mindent el kell követni, hogy az adathalmaz többi része felhasználható legyen. A FOIA arról is rendelkezik, hogy ezen információkat, hacsak az adott intézmény nem teszi amúgy is elérhetővé, írásban kell kérelmezni (mégpedig nem mint információt, hanem mint adat-egységet, a rekord tulajdonságait jól körülhatárolva), és az intézmény ezen kérelmek feldolgozásáért illetve az adatok kiszolgáltatásáért díjat számolhat fel. Ezt a díjat csak abban az esetben köteles elengedni vagy mérsékelni, ha a kérelmező bizonyítani tudja, hogy az igényelt adatok közérdekűek, segítik a kormányzat tevékenységének átláthatóságát. Erre azonban ritkán akad példa.¹⁶⁴

A második megnevezett törvény, a *Privacy Act*, a személyes adatokra vonatkozik, és kettős célt szolgál: egyszerre védi azokat az illetéktelen hozzáféréstől (illetve szabályozza azok megfelelő, méltányos kezelését), másfelől kijelenti, hogy minden egyénnek jogában áll tudni, pontosan milyen adatokat tárolnak róla a törvény hatálya alá eső szövetségi intézményeknél.¹⁶⁵

A harmadik jogszabály (*CIPSEA*) pedig biztosítja a kifejezetten statisztikai célra gyűjtött, egyénekre vagy szervezetekre vonatkozó (bizalmasan kezelendő és ekként bizalmas adatként definiált) információk védelmét. Kijelenti, hogy ez a védelem nem csupán egyéni, hanem közösségi érdek is, hiszen a bizalom megingása csökkenti a statisztikai adatgyűjtés és elemzések illetve ezáltal a döntéshozás hatékonyságát (ezzel az oksági viszonyal magyarázható tehát a törvény kettős célja). Ezért ezeket az információkat a törvény értelmében kizárólag statisztikai célokra lehet felhasználni, illetve nem tehetik lehetővé az alanyok (közvetlen vagy közvetett) beazonosítását.

¹⁶⁴ További információk, illetve a jogszabály magyarázata elérhető a következő Internetes címen (Department of Justice's FOIA homepage): <http://www.usdoj.gov/foia/>, illetve azon belül (FOIA Reference Guide, revised April 2005): <http://www.usdoj.gov/04foia/referenceguidemay99.htm>

¹⁶⁵ A Privacy Act áttekintő magyarázata elérhető a következő helyen: http://www.usdoj.gov/04foia/04_7_1.html

Ennek a törvénynek másik, számunkra különösen érdekes része a statisztikai hatékonyságról szól: hangsúlyozza ennek döntéshozatali jelentőségét (különösen, ami a gazdasági társulásokról gyűjtött információt illeti) és kijelenti, hogy a szövetségi statisztikai hivataloknak folyamatosan fáradozniuk kell hatékonyságuk növelésén, és a nem-válaszolási arány csökkentésén, jóllehet ennek lehetőségét behatárolják az intézmények közötti információáramoltatás törvényben meghatározott korlátai.

Az amerikai rendszer kiterjedt vizsgálata után vegyük sorra a hasonlóságokat és a különbségeket az általunk vizsgált másik két modellhez képest. Nagy-Britanniában, mint említettük, ugyancsak létezik *Freedom of Information Act*, és adatvédelmi törvény is (*Data Protection Act*). Előbbi mutat hasonlóságokat az amerikaival, ám ez nemcsak közhivatalokra terjed ki, némileg más szempontok alapján állapítják meg az árazást, és nem teljesen megegyező azon adatok köre sem, amelyekre már nem terjed ki a törvény. Ez utóbbi ugyan árnyalatnyival szélesebbnek mondható, viszont ebben az esetben külön hangsúlyt kap, hogy a hivataloknál, intézményeknél tárolt összes adatra kiterjed, származzon bárhonnan és bármilyen formában. Érdeemes azonban megjegyezni, bár ez sem jelent lényeges eltérést az amerikai modellhez képest, hogy az adatvédelmi törvény egyértelműen prioritást élvez, tehát csak azok a személyekre vonatkozó adatok igényelhetők, amiket az is lehetővé tesz.¹⁶⁶ Az FOIA előírásainak teljesülését itt információs biztos (*Information Commissioner*) ellenőrzi.

Az északi országokban az alfejezet bevezetőjében említett két szabályozási terület tisztán megjelenik külön adatvédelmi, információszabadság- és statisztikai törvény formájában. Az adatvédelemnél megállapíthatjuk, hogy nem vonatkozik ugyan törvényi korlátozás a beazonosítható adatok kiszolgáltatására, de a gyakorlatban ez mégsem fordul elő. Másrészt, megjelenik az a figyelemre méltó, máshol nem használt kitétel is, hogy a bizalmas adatoknál egy fokkal szűkebb körű érzékeny

¹⁶⁶ A brit információszabadság-törvényről legkönnyebben az információs biztos honlapjáról tájékozódhatunk: <http://www.informationcommissioner.gov.uk/eventual.aspx?id=74> illetve részletes áttekintést itt találunk: <http://www.ico.gov.uk/documentUploads/Freedom%20of%20Information%20Act%202000%20An%20Introduction.pdf>

adatok csak abban az esetben használhatók fel (és akkor is statisztikai vagy kutatási célra, gyakran engedélyhez kötötten), ha a kutatás társadalmi hasznossága nagyobb, mint az egyéni integritás megsértésének kockázata. A bizalmasság értelmezése itt egyébként megfelel annak, ahogyan az amerikai *CIPSEA* a beazonosíthatóságot definiálja, vagyis az északi országokban a beazonosítható adatok minősülnek bizalmasnak.¹⁶⁷

Nem szabad elfelejtenünk, hogy európai országok esetében a EU bizalmas adatok kezelésére vonatkozó szabályozása is mérvadó¹⁶⁸. Ez már továbbmegy a beazonosíthatóság kérdését illetően: ennek megállapításhoz már a racionálisan felhasználható eszközöket is figyelembe kell venni. Végezetül, az adatvédelmi és működési szabályozásokkal kapcsolatban megállapíthatjuk, hogy általában intézményeken belüli szabályozások is vonatkoznak erre, ahogyan ezt a most következő részben is látjuk majd.

8.2.2.A szabályozás szintje és ereje

A szabályozás szintje szerint három szintet különítünk el: intézményi, nemzeti és nemzetek feletti. Ez tulajdonképpen összefügg a harmadik csoportosítással, azaz azzal, hogy egy szabályozás kötelező érvényű-e vagy pedig csak ajánlásként funkcionál. Az előbbiek jellemzően az egyes országok jogalkotásához (illetve intézményen belüli hatállyal belső szabályozásokhoz), utóbbiak pedig a nemzetközi szervezetek standardjaihoz kapcsolhatók, igaz, ajánlásokat nemzeti kormányok is létrehozhatnak.

Az intézményi szabályozás néhol igen jelentős, mint például Angliában az *ONS National Statistics Code of Practice* című statisztikai gyakorlatra vonatkozó szabálygyűjteménye, amiből számunkra különösen érdekes az adathozzáférésről és

¹⁶⁷ Az északi országok vonatkozó jogalkotását illető információink a következő dokumentumból származnak: *Access to Microdata in the Nordic Countries* – az északi országok statisztikai hivatalainak közös kiadványa, Statistics Sweden, SCB-Tryck, Örebro, 2003

¹⁶⁸ *Recent EU legislation for research access to confidential data – implementation and implications*, Conference of European Statisticians, Geneva 10-12 June 2003

bizalmasságról szóló rész (*Protocol on Data Access and Confidentiality*)¹⁶⁹. Mivel itt az ONS, bár nem egyedüli statisztikai intézmény, befolyása az említett összefonódások révén igen nagy, ez a protokoll erősebb hatású, mint általában a belső szabályzatok. Ennek alapján az adatvédelem jelentőségének, illetve a szükséges intézkedések megtételének mérlegelésekor bevonják azt a szempontot is, hogy vajon összekapcsolhatók-e az adatok más táblákkal, de megjelenik az egyének önbeazonosítása is, mint lehetőség. Bár a statisztikai hatékonyság javításának kérdésében nem olyan alapos, mint az amerikai *CIPSEA*, azért ez a szabályozás is kimondja, hogy a minőségre ható adatvédelmi intézkedések megtételét (az adatok preparálást) jelezni kell, metaadat formájában, a többi kiegészítő információval együtt.

Ilyen belső szabályozások, deklarációk az USA-ban is léteznek, hangsúlyos szempont az adatminőség és az adatvédelem is. Utóbbi témában az intézmények rendszerint hangsúlyozzák, hogy mindezt kettős cél szem előtt tartásával teszik: minél több hasznos adatot szolgáltatni a kutatók, és az egész társadalom számára, az egyén integritásának maximális megtartása mellett. A *Census Bureau* például a következő elveket veszi figyelembe a válaszadók érdekeinek védelme érdekében:¹⁷⁰ csak az adott szövetségi program szempontjából valóban szükséges információkat gyűjtik, törekszenek a minél teljesebb körű tájékoztatásra az adatfelvétel megkezdése előtt, figyelembe veszik a válaszadó preferenciáit, igényeit is (hiszen csak akkor van értelme az adatfelvételnek, ha a kérdések túlnyomó részére érkezett válasz), valamint természetesen kiemelt fontosságú a bizalmas adatok védelme is. A *CB*-ban külön szervezeti egység foglalkozik az adatkezeléssel kapcsolatos szabályok, előírások betartásával, ennek alárendelten pedig létezik egy belső, a nyilvánosságra hozást felügyelő, tanács is (*Disclosure Review Board*), ami a bizalmas adatokra vonatkozó elvek betartását ellenőrzi az adatsorok közzététele előtt.

¹⁶⁹ National Statistics Code of Practice – Protocol on Data Access and Confidentiality, megtalálható a következő Internetes címen:
http://www.statistics.gov.uk/about_ns/cop/downloads/prot_data_access_confidentiality.pdf

¹⁷⁰ A *CB*-ra külön törvényi fejezet is vonatkozik: Title 13 United States Code (USC), de intézményi programjuk itt tekinthető meg: http://www.census.gov/privacy/files/data_protection/002822.html

A nemzeti szinttel kapcsolatban meg kell jegyeznünk, hogy az előző fejezetben látott, szerkezetileg igen különböző három modell esetében általános jogalkotási különbségeket is látunk: az elsőben a kontinentális, míg a második kettőben az angolszász gyakorlatot alkalmazzák. Az északi országok vonatkozó szabályozása emiatt valamennyire eltér a másik két modelltől, ahol például igen kiterjedt a kapcsolódó jogszabályok és magyarázatok köre, ám ez elemzésünket nem érinti, mivel tartalmilag igen nagy hasonlóságot mutatnak. Mindegyik esetében alapvető jelentőségű az előző alfejezetben tárgyalt adatvédelmi törvény, ami egyben a legerősebb korlátokat állítja fel a statisztikai adatkezelő intézmények számára.

Nemzetközi szinten annyiban más a helyzet, hogy itt inkább standardokról, irányelvekről van szó. A statisztikai adatokra vonatkozó szabályozás több világszervezetnél és az Európai Uniónál is megjelenik, utóbbinál többértűen is: mind az adatvédelemnek, mind a bizalmas adatokhoz való hozzáférésnek külön fejezetet szenteltek. Ezeket a tanulmány végén, az országos szintű rendszerek tulajdonságaitól elkülönítve taglaljuk majd.

8.3. Adatvédelmi gyakorlat

Az adatvédelem kérdése tehát igen fontos alapelv az általunk vizsgált összes modellben, amely alkotmányos (vagy hasonló jelentőségű) alapjogok garantálásán alapszik, és jellemzően adatvédelmi törvény formájában jelenik meg. Az EU tagjai a vonatkozó közösségi szabályozást is figyelembe veszik. A hangsúly ezek esetében kevésbé az adatfeldolgozás, továbbítás gyakorlati kérdésein van, mint inkább ezen általános jogok biztosításán, amelyeket a lehető legszélesebb körben érvényesíteni kell (tehát például az amerikai *CIPSEA* alapján bizalmas adat nem használható fel nem statisztikai célra, azaz például hatósági intézkedésre).¹⁷¹ Bár megjegyezzük, hogy ez alól akadnak kivételek: gyakran előfordul (például Svédországban), hogy bizonyos állami intézményeknek továbbított adatokra nem ezek a szabályok vonatkoznak, valamint ugyanitt látunk egy érdekes példát arra is, amikor mégis

¹⁷¹ Confidential Information Protection and Statistical Efficiency Act, Title 5 USC, sec. 511 (b) (1), figyelembe véve: sec. 501 (3) (A)

lehetőség van arra, hogy orvosi célú kutatásokra beazonosítható, jellemzően egészségügyi adatokat is kiadjanak.¹⁷²

Ezek az előző fejezetben részletezett kötelező érvényű szabályok és ajánlások egyben kiindulópontul szolgálnak a statisztikai gyakorlat számára, különösen ami a nem közérdekű, tehát egyénekre visszavezethető mikroadatokhoz való hozzáférést illeti. A bizalmas, tehát (természetes vagy jogi) személyekre vonatkozó adatok kezelésére általánosan jellemző ezekben az országokban, hogy hozzáférésük nem nyitott bárki előtt. Az engedélyt leggyakrabban az adatot kezelő statisztikai hivatal vagy intézmény adja meg, néhány szempont vagy egészen részletes kutatási terv alapján, attól függően, hogy milyen mélységben kíván a kutató betekintést nyerni az adatokba. Vannak általános, más úton megszerzett hozzáférési jogosultságok is, mint például Angliában az Athens-azonosító, ezt később részletesebben tárgyaljuk. Az északi országokban ezen kívül létezik egy külön ezt a célt szolgáló adatvédelmi hatóság, ennek az engedélyét jellemzően érzékeny adatokhoz való hozzájutás érdekében kell kérni Dániában, Norvégiában és Svédországban.

Az adatvédelmi szabályozás bizonyos kötelezően megteendő intézkedéseket von maga után, amelyek egyfelől csökkentik az adatok felhasználói rendelkezésre bocsátott adatok információ-tartalmát, így természetesen korlátként jelennek meg számukra; másfelől a szükséges intézkedések megtétele költséget jelent az adatkezelő intézmények számára. Természetesen ennek fejében olyan előnyökkel járnak, amik miatt semmi esetre sem kívánatos ezeknek a korlátoknak a lebontása. Nem sérül az alanyok magánélete, sem pedig az intézmény megítélése, ily módon továbbra is sikerrel folytathat majd bizalmas adatokra épülő kutatásokat.

Ezek az intézkedések általában nem a mikroadatok (tehát az egész, egy egyénre azonos kutatásban vonatkozó összes adatot magában foglaló rekord) továbbadását tiltják, hanem csak bizonyos, az alanyok közvetlen vagy közvetett beazonosítását lehetővé tevő elemeiket emelik ki. Ebben az esetben a statisztikai intézmények kötelesek továbbadás vagy esetleg nyilvánosságra hozás előtt az adatokat

¹⁷² Ld. az északi országok jogalkotásánál már hivatkozott tanulmányt.

preparálni, azaz lehetetlenné tenni az illetéktelen hozzáférést az érzékeny adatahoz. Az így kivett, kitakart információkkal kapcsolatban ezután még több lehetőség is van: előfordulhat, hogy biztonságos körülmények között, külön engedéllyel a kutatók még hozzájuthatnak bizonyos részeikhez (például irányítószámokhoz, mint a PSID esetén), de nem minden részlet érhető így el. Több helyen kifejezetten hangsúlyos az érzékeny adatok fizikailag is biztonságos tárolása, amit úgy is biztosítanak, hogy például az így, külön engedéllyel nyert adatokhoz csak az adatkezelő intézmény adott helyiségeiben lehet hozzájutni, és ott is csak úgy, hogy a munka eredményét be kell mutatni az intézmény egy alkalmazottjának.

Több módszer ismeretes az adatok biztonságosabbá tételére. Lehetséges, hogy egyszerűen eltávolítják a közvetlen azonosítókat (név, cím, telefonszám, személyi szám, stb.) illetve ahol általános numerikus kódokat használnak (mint például az északi országokban, ahol épp ezek által válnak összekapcsolhatóvá az adatállományok), ott azt inkább ideiglenesen egy másik kóddal helyettesítik. De földrajzi korlátokat is felállíthatnak, tehát például úgy, hogy bizonyos elemszámnál kevesebbet tartalmazó körzetek azonosító jelét eltávolítják. Ehhez hasonló az a módszer, amikor bizonyos változók kategóriáit átszabják, összevonják (igen gyakori ez például iparágaknál, ahol a pontos megjelölést egy tágabbal helyettesítik), vagy az értékeit kerekítik (például jövedelemnél lehet ezt megtenni), osztályközökbe sorolják (ezt pedig leggyakrabban életkornál alkalmazzák).

Előfordul olyan eset is, hogy úgy változtatják meg az adatokat, hogy az egyes elemekre lebontva ugyan már nem, de sokasági tulajdonságaira nézve azonosak maradnak. Ilyen módszer az is, amikor bizonyos kiugró (tehát valamilyen más, különleges tulajdonságuk alapján beazonosítható) adatokat kitakarnak, de előfordul radikálisabb változtatás is. Például amikor bizonyos tulajdonságaik alapján ugyanolyan, de földrajzilag máshol elhelyezkedő egységeket felcserélnek, vagy, ilyen módszer a szimuláció, ami a legkevésbé „valóság-hűnek” mondható. Ezt főleg olyankor alkalmazzák, amikor az adott adatsor iránt nagy a kereslet, viszont jellemzően nem részletekbe menő kutatást végeznek, ezért megfelelő lehet egy

nem valós, de a valódival megegyező sokasági tulajdonságokkal (tehát például eloszlással, átlaggal, rétegződéssel) rendelkező sor felhasználása is.

Az, hogy szükség van-e ilyen módszerek alkalmazására, és főleg, hogy melyikre, az függ az adatsor típusától, attól, hogy hány változót tartalmaz, és természetesen a bevett gyakorlattól is. A különböző módszerek másképp módosítják az adatokat, ezért az adatminőség szempontjából fontos információ lehet, hogy mely(ek)et alkalmazták. Mint láttuk, néhol (például Angliában, az *ONS*-nél) az adatokat előállító intézmény köteles az elvégzett módosításokról az adatsorhoz mellékelni (vagy külön letölthető) metaadatok között számot adni. Természetesen, az adatminőség minden általunk vizsgált országban fontos szempont, valamint általában kifejezetten törekednek azok jó felhasználhatóságára, így a változtatásokról is számot adnak.

Az adatvédelemhez szorosan kapcsolódik a hozzáférés a következő alfejezetben tárgyalt kérdése, hiszen az adatok illetéktelen kézbe kerülését szolgálja alapvetően az is, hogy csak bizonyos, jól körülhatárolható felhasználó körök számára lehetséges a hozzáférés. Ezzel bővebben a következő fejezetben foglalkozunk, itt általánosságban annyit érdemes megemlíteni, hogy széles nyilvánosság elé gyakorlatilag egyáltalán nem kerülnek mikroadatok, mivel ezek feldolgozatlanok, kezelni, értelmezni is tudni kell azokat, és az adatgazda intézmények általában ügyelnek arra, hogy elkerüljék a félreértelmezés lehetőségét. Ezt néhol (például Angliában) központilag támogatott kurzusokkal, tanfolyamokkal is igyekeznek elérni, de azokat is csak olyanok számára szervezik, akiknek eleve van hozzáférési jogosultságuk. Ezzel az adatok mellett a szakmát is védik természetesen, hiszen, az adatgyűjtő intézmények számára elemi fontosságú a széles közvélemény bizalma.

8.4. Hozzáférés az adatokhoz

Az adatokhoz való hozzáférés (tehát, mint a jogszabályokkal kapcsolatban megállapítottuk, nem jogi, hanem gyakorlati értelemben) részben az intézményi struktúra által is meghatározottnak tűnik. Azt látjuk, hogy a legszigorúbb feltételek az északi országokban kapcsolódnak a mikroadatok eléréséhez, ott a

legkörülményesebb ez a folyamat, ahol a leginkább központosított is. Igaz, ezek az országok dolgoznak egyben (legalábbis a vizsgálatunk körén belül) a legkisebb adathalmazzal, és talán az is igaz, hogy számszerűen kevesebben érdeklődnek a mikroadatok iránt. Az adatokhoz való hozzáférést is több oldalról érdemes megközelíteni: milyen fizikai lehetőségei vannak, milyen körben teszik elérhetővé, van-e valamilyen elengedhetetlen feltétele a hozzáférésnek, illetve milyen minőségűek, mennyire felhasználhatóak az így nyert adatok.

A hozzáférés fizikai módja szerint négyféle módszert alkalmaznak jellemzően. *Off-site* elérés esetén valamilyen adathordozón (CD-n vagy mágneslemezen) kiküldik az adatot a statisztikai hivatalon kívülre, és a felhasználó saját környezetében dolgozhat rajtuk. Ekkor a kutatás közben nincs sem virtuális, sem fizikai kapcsolat az intézmény és a felhasználó között. Az *on-site* elérés már fizikai kapcsolatot jelent, ekkor a hozzáférést kutatószobában teszik lehetővé, ami a statisztikai hivatal épületeiben található. *Off-line* és *on-line* elérésről beszélünk, amikor a felhasználó az Interneten keresztül lép kapcsolatba az adatszolgáltatóval. A különbség az, hogy előbbi esetben a felhasználó egyáltalán nem látja az adatokat (csak megadja a műveleteket, amiket az adattáblán végre kíván hajtani), utóbbinál viszont jellemzően egy megfelelően biztonságos kapcsolat létrehozása után letöltheti a rendelkezésére bocsátott adatokat (ezek általában egyedi azonosítók, Angliában pedig eleve intézményi tagsághoz kötött, hogy valaki rendelkezik-e ezzel a biztonságot bizonyító státusszal, vagy nem).

Az általunk vizsgált modellekben viszonylag elterjedt az on-line elérés, bár az északi országokban egyelőre csak Dániában, illetve szűkebb körben Svédországban alkalmazzák. Az USA-ban már több felvételnél lehetséges ez, például a CPS, az NLS, a PSID vagy a HRS adatai is letölthetőek, de ott még számos más típusú hozzáférést is találunk, főleg bizalmas adatokhoz. Ahol pedig létezik már az on-line mód, ott bár néhol az elérés kevésbé egyértelmű, hasznos szolgáltatásként adatbányász eszközök is használhatók. Ennek segítségével a felhasználó könnyebben, még a letöltés előtt tájékozódhat az adatok pontos tartalmáról, formátumáról, használhatóságáról.

Angliában mára igen nagy mennyiségű adatot tettek ily módon elérhetővé, a UKDA-n keresztül. Elérhetőek itt anonimizált mikro- és makro adatok, többféle formátumban, de akár egyes adatfelvételek kvalitatív részei is, mint például mélyinterjúk, naplók, stb. Az adatbázis piackutatásból, történeti kutatásokból származó adatokat is tartalmaz, ezekhez is ugyanolyan módon lehet hozzájutni, mint bármely másikhöz. A hozzáférés a brit akadémiai életben általánosan használt, központosított Athens jogosultsággal lehetséges. Ez főleg brit felsőoktatási és továbbképzési intézmények hallgatói és dolgozói számára elérhető, de akár külföldi intézmények tagjai is igényelhetik, igaz, csak szűkebb jogosultságot kaphatnak (például nem vonatkozik a népszámlálás adataira, ugyanis ott más rendszer szerint adják ki az adatokat). Az elérésre jogosult felhasználók (és ebben a nagyfokú állami támogatottságnak is van szerepe) a legtöbb adatállományhoz ingyen juthatnak hozzá, amennyiben kutatási, akadémiai vagy oktatási célra használják ezeket. Üzleti célú felhasználásért meglehetősen nagy költséget számolnak fel.

Sok helyen, főleg az északi országoknál vagy Anglia esetében (bár itt többnyire kiegészítő jelleggel) találkozunk off-site eléréssel. Ilyenkor az adatokért mindenképpen fizetni kell, viszont előfordul, hogy így jobb minőségű, esetleg különböző segédanyagokat tartalmazó adatbázishoz jutunk. Ennek a formának megvannak az előnyei, gyakran jobban használható formátummal, több tájékoztatóval adják ki őket, és sokszor könnyebb is ilyen formátumban kezelni. Utóbbi okozhatja, hogy olyan példát is találunk, ahol az ingyenes letölthető, és a költségtérítéses CD-n található adatbázis pontosan ugyanazt tartalmazza, mégis van igény mindkettőre. Az intézmény számára egyébként ennek a módnak vannak biztonsági előnyei is, hiszen Internetes kapcsolatnál mindig fennáll az illetéktelen behatolás veszélye, így viszont a hivatal biztos lehet benne, hogy a felhasználó csak azt kapja, amit megrendelt. Viszont kétségtávol vannak hátrányai, ugyanis meglehetősen magas a költsége a felhasználó számára: egyrészt ki kell fizetni a termék árát (erről részletesebben lejjebb), de költségként fogható fel az a néhány nap, esetleg még hosszabb idő, ami alatt az adat eljut hozzá. Ezen felül az a tulajdonság is fontos lehet, hogy ha nincsen on-line módja az adatlekérésnek, akkor

előfordulhat, hogy a kutató nem tudja pontosan megítélni, melyik adattábla mennyire használható számára, így esetleg rosszat rendel.

Sok helyen alkalmazzák az on-site elérést is, általában érzékeny adatok esetében. Igen elterjedt például az Egyesült Államokban, ahol bizonyos mélységű adatokat, vagy olyanokat, amiket másképp nem tesznek közzé, az úgynevezett Data Centerekben lehet elérni, szerte az Államokban. Ezek a CB „fiókszervezeteiként” működnek, céljuk az adatszolgáltatás állami és nem állami intézmények, szervezetek és magánszemélyek részére, ingyen, vagy egyénre szabott szolgáltatások esetében költségtérítéssel. Hasonlóan, és tulajdonképpen a bizalmas adatok iránti stabil igény miatt mondhatjuk úgy is: természetesen a BLS is tart fenn ilyen központot, bár csak egyetlen helyen az Államokban. Így tehát közvetlenül a statisztikai adatszolgáltatónál a felhasználó több, számára esetleg lényeges, ám egyébként kitakart változóba nyerhet betekintést, ám fokozott ellenőrzés, és nyilván, helyhez kötöttség mellett.

Különbséget tehetünk aszerint is, hogy milyen felhasználói csoportok számára engedélyezett a hozzáférés. Az adatkezelők alkalmazhatnak tiltást bizonyos körrel szemben (például az üzleti felhasználók esetében több skandináv országban), megszabhatnak egy kizárólagos hozzáférési módot (például Dániában külföldiek számára csak a koppenhágai on-site elérés lehetséges)¹⁷³ vagy egyszerűen előírhatnak körülményes engedélyeztetési procedúrát. Erre a legjobban megfigyelhető példákat az USA-ban találjuk: a CB bizalmas adataihoz például úgynevezett Special Sworn Status szükséges,¹⁷⁴ ami tulajdonképpen egy írásos megegyezés, és amihez kutatási tervet kell bemutatni, amit tudományossági, megbízhatósági és tárgyi szempontok alapján vizsgál meg a CB illetve annak Internal Revenue Service nevű bizottsága, de külföldiekre hasonló folyamat vár a BLS-nél is. Ezek mellett alkalmazhatnak diszkriminációt az árszabásban, ami szintén elkülönít bizonyos csoportokat, vagy állíthat belépési korlátokat.

¹⁷³ Ld. Az idézett tanulmányt az északi országokkal kapcsolatban (23. oldal).

¹⁷⁴ Erre látunk példát a Michigan Census Research Data Center esetében (<http://www.isr.umich.edu/src/mcrdc/more.html>), illetve törvényi úton is szabályozott: Title 13 USC, 23 (c)

A legjellemzőbb bontás egyébként a jogosultság körének csökkenő sorrendjében: az intézményen belüli kutatók, intézményen kívüli kutatók (és diákok), külföldiek, más hivatalok, nyilvánosság. Az egyes csoportok között a hozzáférés gyakran az elérhető adatok körében is különbözik egymástól. Például a széles nyilvánosság elé gyakorlatilag egyáltalán nem kerülnek mikroadatok, míg bizonyos, nagy elismertséggel rendelkező intézetek, egyetemek kutatói előtt jóval kisebbek az akadályok. De megfogalmazhatjuk a felhasználói kör egy másik felosztását is, például Angliában az első két csoport helyett mindenütt a tudományos és üzleti tevékenység megkülönböztetését emelik ki, tehát a hangsúlyok a felhasználó szakmai, intézményi háttere felől a felhasználás célja felé tolódnak, másképp megfogalmazva: az intézményi protekcionizmus felől a piaci alapú megkülönböztetésre.

Az elérhető adatok felhasználhatósága szerint is fontos különbséget tenni, hiszen mint láttuk az off-site elérés esetén is, gyakran előfordul, hogy egy adatsorhoz akár több helyről vagy több formátumban is hozzá lehet jutni. Ilyet találunk például az Egyesült Államokban, ahol a CPS adatai elérhetők (letölthetők) az NBER-on keresztül, vagy a BLS-en keresztül, a CPS felvétel honlapján. Előbbi helyen részletesebb leírás, felhasználóbarát segédprogramok, anyagok segítik az adatok iránt érdeklődőt, utóbbi helyről pedig többféle formátum érhető el. Példaként megemlíthető az angol népszámlálás adatainak elérhetősége is, ami ugyan alapvetően egy helyről, a UKDA-n keresztül lehetséges, és akkor sem a teljes állományra, csak a teljes népszámlálási adathalmazból vett 1-5% minták formájában (Samples of Anonimized Records - SAR). Ez tulajdonképpen hasonló ahhoz, mintha eleve mintavételes eljárással készült felmérés adatai lennének, ám a mintaelemszám sokkal nagyobb, ami az elemzést a kisebb földrajzi egységek, illetve kisebb népességi csoportok (etnikumok, vallási kisebbségek) szintjén is lehetővé teszi.

Hasonló korlátozást látunk Dániában is, ahol ugyan létezik már on-line elérés (igaz, csak megfelelő jogosultság esetén), ám az így hozzáférhető adatok minden esetben

az eredeti sokaságból vett minták rekordjai. Ennek oka valószínűleg hasonló az angol népszámláláshoz, nevezetesen hogy itt is kiterjedt regiszterekről van szó.

8.5. Árazás

Az árazás során az egyes intézmények az általános részben leírt stratégiákat választhatják, ezekre mutatunk be példákat a vizsgált modellekben. A követett gyakorlatot természetesen meghatározza az intézmény állami beágyazottsága, alapvetően maga a finanszírozás módja, de azt látjuk, hogy lényeges tényező az intézményi hagyomány és a társadalmi környezet is.

A szolgáltatások előállításánál felmerülő költségeket minden intézménynek fedeznie kell valahogyan. Mivel az adatok természetéből (és például egyes intézmények nyilvántartási funkciójából) adódóan a fizetőképes kereslet mindig igen kicsi lesz az elvégzendő feladatokhoz képest, így mindegyik intézmény igénybe vesz állami támogatásokat. Ennek nagysága függvényében azután bizonyos költségeket az adatok felhasználóra hárítanak. Azt látjuk, hogy a statisztikai adatfeldolgozással foglalkozó intézmények majdnem mindegyike valamilyen állami, kormányzati szervhez is szorosabban kötődik (mint az amerikaiak közül a *BLS*, Angliában az *ONS*, és természetesen, az északi országok valamennyi hivatala), de vannak egyetemi kutatóintézetek is (a *PSID* a Michigani Egyetemen készül, vagy az *ISER* is ilyen). Viszonylag ritka ezen a szintén a tisztán magán intézmény, aminek alighanem az is oka, hogy egy komolyabb adatfelvétel majd az összegyűjtött adatok kezelése igen nagy apparátust igényel, és igen rossz közvetlen megtérülést ígér.

Az állami támogatás arányát azonban más tényezők is befolyásolhatják. Például Angliában, ahol viszonylag decentralizált rendszerrel találkozunk, különösen nagy összegű kormányzati támogatásokra számíthatnak ezek az intézetek, aminek valószínűleg társadalmi hozzáállásbeli okai vannak. A magas kutatási támogatások és a törekvés arra, hogy egy különösen hatékony, felhasználóbarát hozzáférési rendszert hozzanak létre (lásd az *ESDS* együttműködés és persze a *UKDA* tevékenysége), arra is utal, hogy a kormányzat számára lényeges

információtartalommal bírnak az empirikus kutatások eredményei, amik visszacsatolást biztosítanak egyrészt a társadalom felől, másrészt pedig legitimációs erővel is bírnak. Érdekes itt megemlíteni azt is, hogy ez az elem az amerikai kutatásban is megjelenik, például az SIPP nevű felvétel is a visszacsatolást, a kormányzat elsődleges információszerzését szolgálja.

Az árképzésben számos diszkriminációs elvet figyelhetünk meg. Más az ára természetesen a különböző elérési szolgáltatásoknak, de más lehet a felhasználótól függően is: földrajzi elhelyezkedése (itt nem hagyhatjuk figyelmen kívül a fizikai elérések utazási, várakozási költségeit), intézményi hovatartozása, felhasználási célja, megbízhatósága mind befolyásolják az adatszolgáltatás árát.

Általánosságban igaz, hogy az on-site elérésért mindenütt fizetni kell, és általában egyfajta önköltséget számolnak fel. Az északi országokban ezt úgy fogalmazzák meg, hogy az adatok előállítási költségeit a felhasználó fizeti, ugyanis az ár képzése a teljes költségre vetítve történik (közvetlen munkaköltség, ingatlan bérleti díj, irodai és személyzeti költség, marketing és az irányítási költségek egy része...), az így számított önköltséget pedig végül hozzáadják az adott csoport dolgozóinak díjához.¹⁷⁵ Ez az USA-ban is hasonló módon történik, a CB szenzitív illetve bizalmas adataihoz hozzáférést biztosító *Research Data Centerekben*.¹⁷⁶ Itt az elérés hosszadalmas kérelmezés után, felügyelettel és utólagos ellenőrzéssel történik, és ezekért munkadíjat számolnak fel, amiből a kutatóközpont működését fedezik. Ez egyébként ellentétes azzal az USA-ban általánosan bevett gyakorlattal, hogy az adatok on-line elérése ingyenes.

Az árdiskrimináció vizsgálatakor tehát, már csak ezen utóbbi megállapítás miatt is, a skandináv és az angliai gyakorlatot érdemes összehasonlítani. Angliában, mint már láttuk, létezik egy erős felhasználási cél szerinti megkülönböztetés. A legtöbb adatállomány ingyen letölthető, amennyiben kutatási, akadémiai (tehát akár oktatási) célra használják ezeket. Üzleti célú felhasználásért meglehetősen nagy költséget számolnak fel. Részletesebben ez a következőképpen néz ki: nem üzleti

¹⁷⁵ Ld. Az idézett tanulmányt az északi országokkal kapcsolatban (16-17.oldal).

¹⁷⁶ Ld. Ezzel kapcsolatban a már hivatkozott MCRDC honlapot.

célú felhasználáskor csak a megrendelhető CD-ért kell fizetni, az árban külön kalkulációs része van a CD előállítási költségének (fix költség) és a fájlhasználatnak (ez volna a változó rész, de valójában a CD-használat határkölsége 0), ami itt még meglehetősen alacsony. Nem úgy, mint az üzleti célú fájlhasználati díj, ami £500 (plusz áfa) és ezt újra be kell fizetni, amikor új célra használják ugyanazt az adatsort.¹⁷⁷ Ez az ár tehát már igen messze van attól az optimális kétrészes árazástól, amikor a belépési költség után csak határkölség alapú kalkulációt alkalmaznak. Itt inkább a piaci érték elve érvényesül, ez látszik abban is, hogy nyilvános adatokért (természetesen) nem kell fizetni.

Az angol diszkriminációs rendszer egyébként jó példa a licenz-típusú árazásra is, különösen az angol népszámlálási adatoknál, amik hozzáférése jobban részletezett. Ott például az is ki van kötve, hogy ezért a fájlhasználati díj 10 ember hozzáférését engedi, további személyekért £50 kiegészítő díjat kell fizetni. Ezért a díjért azonban a hozzáférő alkalmazottak jogosultak részt venni a *Centre for Census & Survey Research (CCSR)* által szervezett tréningeken, illetve segítségért is ide fordulhatnak. A már egyszer lekért adattáblákat pedig fel lehet használni más kutatásokban, illetve más kutatók által is, ám ezeket be kell jelenteni, illetve minden új felhasználónak regisztrálnia kell¹⁷⁸.

Az angliaitól némiképp eltér az északi országok árazási gyakorlata, ahol ezen a téren is jobban tükröződik az elosztó állam, mint központi szereplő jelentősége. Dániában például az állami szabályozás kimondja, hogy ennek a tevékenységnek nullszaldósnak kell lennie, épp ezért nem ingyenes. Itt a szolgáltatások között tesznek különbséget: az államilag támogatott, hatósági regiszterek használatára csak egy, az önköltséget is magában foglaló óradíjat szabnak ki, a szolgáltatások regiszteréért és más, hatóság által fenntartott adatokért pedig extra díjat is felszámolnak, amit a jövőbeli kereslet alapján határoznak meg. A díj (főleg on-site elérésnél) az alkalmazottak szakmai fokozatától függően is változik, de egységárat is használnak. Finnországban pedig, ahol csak off-site és on-site elérési lehetőség

¹⁷⁷ A UKDA árazásáról a következő Internetes oldalról tájékozódhatunk: <http://www.data-archive.ac.uk/orderingdata/charges.asp>

¹⁷⁸ A brit népszámlálás adatainak lekéréséről a következő Internetes oldalról indulva szerezhetünk információkat: <http://www.ccsr.ac.uk/sars/>

van, piaci alapú árazást alkalmaznak, kivéve néhány ingyenes szolgáltatást. Ez azt jelenti, hogy a tevékenység a fogyasztók által finanszírozott, az önköltségben benne foglaltatnak itt is a bérleti díjak és a kutatói helyek felszerelésének költségei, tanácsadó szolgálat, számítási költségek, licenz, stb.¹⁷⁹

Az árazáshoz kapcsolódóan mindenképpen szót kell ejtenünk arról is, milyen egyéb költségekkel jár az adatokhoz való hozzáférés. Számos ilyen költségelemet felsoroltunk már, de nem árt látni, hogy ezek az explicit árakon túl milyen további korlátokat állítanak a felhasználó elé. Ilyen maga a hozzáférés köre, a megfelelő jogosultság megszerzése: ez gyakran lehetetlen, más megfogalmazásban végtelen költségű, például Angliában a népszámlálás adatainak bizonyos állományára nézve. De hasonlóan magas költsége lehet annak is, ha valamilyen adat csak egyetlen központban hozzáférhető, mint például Dánia esetében, ahol a külföldi állampolgárok kizárólag a koppenhágai központban férhetnek hozzá bármilyen adathoz. Ilyen esetekben nyilván az utazási költség is komoly tényező. Más helyeken lehetőség van kiválasztani a hozzáférés módját: kérhetünk szűkebb körű adatot kevesebb pénzért vagy ingyen, esetleg megrendelhetjük valamivel többért, de kiegészítő információkkal, illetve választhatjuk a legköltségesebb, ám legszélesebb körű adathalmazhoz hozzáférést adó on-site elérést.

8.6. Társadalmi környezet: a közvélemény és a kommunikáció szerepe

Az adatszolgáltató intézmények jelentős mértékű interakcióban állnak a társadalom egyéb, nem üzleti szféráival, mint a törvényhozással, (áttételesen) a politikával, közvéleménnyel, a tudományos élettel. Egyszerre kell tehát ezeknek, valamint gazdasági érdekeiknek is megfelelniük működésükkel. Az, hogy ezek milyen hangsúllyal szerepelnek, leginkább országok szintjén különböztethető meg, inkább a társadalmi tényezőkkel, mint a piaci szerkezettel összekapcsolva.

Ezeknél a szempontoknál jellemző, hogy belső, intézményi szabályok adják meg a kereteket, nem pedig állami intézkedések. Ennek oka lehet, hogy a közvélemény

¹⁷⁹ Ld. Az idézett tanulmányt az északi országokkal kapcsolatban (16-17. oldal).

szerpe ebben a rendszerben egy intézmény megítélését jelenti, ami nyilvánvalóan az intézménynek magának is érdeke. Az intézmények körében kiemelkedő szerepe van így az adatvédelemnek és az adatminőségnek. Előbbi az alanyok, a szélesebb nyilvánosság számára jelent pozitívumot, utóbbi a felhasználók számára.

A közvélemény szerepének legrészletesebb kibontását a kutatás során az északi országok esetében találtuk, ahol nem az intézményi, hanem inkább a társadalmi szempont kerül előtérbe. A témával foglalkozó szervek és intézmények például megállapították, hogy bár növekvő igény mutatkozik az adatok kutatási célú kiszolgáltatására, mérlegelni kell, hogy ezen igény mellett hogyan védhetők meg az egyénekre vonatkozó bizalmas adatok, hiszen ha ez a védelem nem elégséges, az politikai vihart is kavarhat. Ettől függetlenül csak viszonylag kevés közvélemény-kutatási adattal rendelkeznek arról, hogy a közvélemény valójában támogatja-e az adatok kutatási célú felhasználását, vagy nem. Norvégiában például általában fontosnak tartják az adatok integritását, és hogy az emberek tudják, milyen információk kérhetőek le róluk. Tanulmányozták a média szerepét is, hiszen a média egyszerre érdekelt a statisztikai adatok bizonyos szintű ismeretében (maga is felhasználja), és a konfliktusok kiélezésében is. Megállapították, hogy a médiafigyelem elsősorban a kereskedelmi célú felhasználásra, ennek kapcsán pedig az egyének feletti ellenőrzésre irányul. Hasonló példa, hogy Svédországban minden második évben készítenek felmérést, és eszerint az embereknek csak kis hányada tartja kockázatosnak, hogy a statisztikai adatait nem kívánt módon is felhasználhatják.

Ehhez képest az angolszász országoknál inkább az egyes intézmények szintjére tolódik a kérdés. A jogalkotásnál láttuk, hogy állami szinten is megjelenik ez az aspektus: az USA-ban a törvény (*CIPSEA*) szövege, Anglia esetében pedig például az *ONS* működési szabályzata is hangsúlyozza, hogy az intézményekkel szembeni bizalom az intézmény hatékony működésének egyik alapja. Ugyanakkor mégis azt mondhatjuk, hogy magának a statisztikai intézménynek, mint szolgáltatónak a megítélését illetően ezek az országok inkább egyfajta piaci szemléletet képviselnek. A társadalom az adatszolgáltatási gyakorlat során már mint kereslet jelenik meg, a

reputáció pedig mint a termék (a statisztikai adatfeldolgozói szolgáltatás) értékét növelő tényező.

Ennek megfelelően ezek az intézmények fokozott figyelmet fordítanak a felhasználóknak nyújtott egyéb szolgáltatásokra. Minden adatállományhoz található egy részletes használati utasítás, elérhető a kérdőív és más, az adat-előállításához kapcsolódó metaadatok, valamint sok helyen letöltés vagy megvásárlás előtt lehetőség van az adattáblák szerkezetének megnézésére, előzetes tájékozódásra (például az angol népszámlálás SAR-adatainál a Nesstar nevű eszközzel). Látunk arra is példákat, hogy akár még az adatfeldolgozást segítő tréningeket, konferenciákat is szerveznek¹⁸⁰ - ezzel egyrészt támogatják a szakmai életet, másrészt az intézmények így javítják saját megítélésüket is. Hasonlóan a felhasználók számára lényeges információt nyújtanak azzal is, ha minőségi nyilatkozatot tesznek, részletesen leírják, milyen korrekciókat végeztek, pontosan milyen módszertant alkalmaznak az adatok elsődleges feldolgozása során, mi az, amit elérhetővé tesznek.

Azt látjuk tehát, hogy ezek az információk több célt szolgálnak. Egyrészt elengedhetetlenek az intézmények számára, de ugyanennyire fontosak társadalmi, kormányzati szempontból is. Mint láttuk, ennek jelentősége a törvények szövegében is megjelenik, ám nem elhanyagolható szempont az sem, hogy a statisztikai rendszer transzparenciája önmagában is javíthatja a kormányzati munka átláthatóságát, ily módon növelve annak legitimációját. Erre következtethetünk abból, hogy Nagy-Britanniában igen jelentős állami részvétel és támogatás mellett működnek (igen színvonalasan) ezek az intézmények.

8.7. Nemzetközi szervezetek adatszolgáltatási gyakorlata

Ebben az alfejezetben két szupranacionális szervezet, az IMF (Nemzetközi Valutaalap) és az Európai Unió adatszolgáltatással kapcsolatos szabályalkotását vizsgáljuk. Előbbi esetében, bár a nemzetközi piacokon igen jelentős gazdasági

¹⁸⁰ Ld. a következő oldalt: <http://www.ccsr.ac.uk/sars/events/index.html>

szervezetről van szó, a mikroadatokkal való rendelkezés nem jelenik meg elkülönülten, inkább a tagországok makrogazdasági mutatóin van a hangsúly, ezért az IMF standardjairól csak rövid áttekintést adunk. Az Európai Unió ennél kiterjedtebben foglalkozik statisztikai adatokkal, minthogy a közösségi statisztikákat az Eurostat állítja elő és kezeli. Így ezen a területen elkerülhetetlen volt a szabályalkotás, mára az adatok kezelésével kapcsolatos irányelvek, jogszabályok is több évtizedes történettel rendelkeznek. Ez utóbbit ezért némileg részletesebben taglaljuk.

8.7.1. Az IMF szabályozása

Az IMF-nél (ahogyan ez általában is hangsúlyos) külön szempontokat vesznek figyelembe a fejlett és a fejlődő országok statisztikáinak minőségi értékelésekor, ezért eszerint különböztethető meg a két egyezmény, és az országoknak lehetőségük van (nem kötelező) valamelyik aláírására.¹⁸¹ A fejlődő országok számára készült az úgynevezett *General Data Dissemination System*, ami az adatszolgáltatás alapvető javítását célozza, és igyekszik ehhez támogatást is biztosítani, a másik pedig a globális pénz-és tőkepiacokon már fejlettnak számító országok számára készült *Special Data Dissemination Standard*, ez már feltételez bizonyos kialakult statisztikai intézményrendszert, és inkább az ellenőrzésben van szerepe. A kettő között természetesen fokozatosság van: ha egy fejlődő ország már elért egy megbízható adatszolgáltatói színvonalat, továbbléphet a második szintre. Korábban arra is volt lehetőség, hogy egy ország csak fokozatosan alkalmazza saját rendszerére az IMF elvárásait, mára ez megszűnt, stabilan kétfokozatú a szabályozás. Ezek mellett az IMF külön fórumot tart fenn a minőség fejlesztésének, az erről szóló párbeszéd elősegítésére (*Data Quality Reference Site*). Érdekes tanulmányokat találunk itt, mint például az, ami a Svájci Szövetségi Statisztikai Hivatal (*peer-review*) értékeléséről szól. Ez első ismert példája annak, hogy egy nemzeti statisztikai hivatal önként kérte külső megfigyelők véleményét. Az elemzés

¹⁸¹ Az ide vonatkozó dokumentumok elérhetők az ezt közzé tevő szervén, az *IMF Dissemination Standards Bulletin Board*-on keresztül: <http://dsbb.imf.org/Applications/web/dsbbhome/>

a statisztikai rendszer alkalmazkodóképességét, hatékonyságát és szavahihetőségét vizsgálta. A továbbiakban rövid, tárgyszerű leírást adunk a két standardról.

8.7.1.1.Special Data Dissemination Standard

Azon tagok számára adták ki, akik jelen vannak, vagy belépni szándékoznak a nemzetközi tőkepiacokra, ennek megfelelően a gazdasági és pénzügyi (makro) adatokkal való rendelkezést taglalja. Aláírása önkéntes (Magyarország is aláírta), amivel a tagok elkötelezik magukat a standardnak való megfelelés folyamatos ellenőrzése mellett és rendszeres tájékoztatást adnak az IMF-nek adatszolgáltatási gyakorlatukról. Alapvetően gazdaságpolitikai célja van: az adatok elérhetőségének javítása növeli a hatékony gazdaságpolitika lehetőségeit, és persze csökkenti a különben könnyen felmerülő inkompatibilitási problémák megoldásának költségét is. Hangsúlyos az adatok integritása (kormányzati korrigálásokat igyekeznek kiszűrni) és a nyilvánosságra hozatal módja. Az aláíróknak metaadatokat is közölniük kell, amiket az IMF elektronikus úton tesz közzé. A szervezet törekvései közé tartozik bizonyos adatok és metaadatok Interneten való cseréjének biztosítása, más gazdasági világszervezetekkel együtt részt vesz a Statistical Data and Metadata Exchange elnevezésű, egyelőre kezdeményezés-szintű projektben.

8.7.1.2.General Data Dissemination System

Ennek aláírói jellemzően fejlődő országok. Céljai az adatminőség és adatszolgáltatás javítása, fejlesztési tervek készítése, irányadás a hatékony makrogazdasági politikára. Kezdetben csak gazdasági, pénzügyi adatokról szólt, 2001-től társadalmi, demográfiai adatokra is kiterjed. Az aláíró országok ország-kooordinátort jelölnek ki és folyamatosan dokumentálják statisztikai adatelőállítási folyamatukat. A hangsúly inkább a minőség folyamatos javításán van, mintsem határidőre elvégzendő feladatokon. Itt is lényeges az adatok esetében azok integritása és a pontos nyilvánosságra hozatal, a hozzáférés gördülékenysége.

8.7.2. Az Európai Unió szabályozása a bizalmas adatokhoz való hozzáférésről és azok kutatásáról

Statisztikai adatokról szóló szabályozást már az Európai Közösség alapító okirata is tartalmaz. Az adatok előállítása a következő alapelvek figyelembevételével történhet: pártatlanság, megbízhatóság, tudományos függetlenség, költséghatékonyság és statisztikai bizalmasság. Az Eurostat által kezelt adatokra azután további rendelkezések születtek, az egyik legfontosabb az 1997-es „Statisztikai törvény” (1997/322). Ebben a bizalmas adat definíciója (13. cikk):

„A nemzeti hivatalok illetve a Közösség hivatala által használt, közösségi statisztikák előállítására szolgáló adatok akkor tekintendők bizalmasnak, amennyiben lehetővé teszi a statisztikai alanyok beazonosítását, akár közvetlen, akár közvetett módon azáltal, hogy egyéni információt tartalmaznak. Annak meghatározásához, hogy egy statisztikai alany beazonosítható-e, tekintetbe kell venni az összes olyan eszközt, ami egy harmadik fél által *az ésszerűség határain belül felhasználható* egy statisztikai alany beazonosításához.”¹⁸²

A törvény emellett megállapítja, hogy nem minősülnek bizalmasnak a nyilvánosság számára elérhető forrásokból származó adatok. A Statisztikai törvény azt is előírja, hogy bizalmas adat csak statisztikai célokra használható fel, kivéve ha a válaszadók félreérthetetlen beleegyezésüket adták adataik más célokra történő felhasználására, illetve külön kitétel vonatkozik a tudományos kutatásra. (Így például valamennyi nemzeti statisztikai hivatal hozzájárulásával, kutatási célokra egyszerűen hozzáférhetővé vált az Európai Közösség háztartási panelfelvételének (ECHP) anonimizált adatállománya)

Az anonimizált adatot a szabályozás a következőképpen határozza meg: „olyan egyéni statisztikai rekordok, amiket úgy módosítottak, hogy a mindenkor legjobb gyakorlat szerint a lehető legkisebbre csökkentsék azon statisztikai egységek beazonosíthatóságát, amikre vonatkoznak”.¹⁸³

¹⁸² Saját fordítás, ezen belül saját kiemelés.

¹⁸³ Saját fordítás

A 2002-es új szabályozás megalkotásakor az Eurostatnál már felismerték a mikroadatok jelentőségét, és igyekeztek megkönnyíteni a hozzáférést. Az Eurostat egy tanulmánya¹⁸⁴ részletesebben is taglalja, miben is áll ez a változás, ami miatt elkerülhetetlen a mikroadatokra vonatkozó szabályozás újragondolása. Egyrészt napjainkra több adat szükséges a kormányzat elszámoltathatóságához, az átláthatóság teljesüléséhez, illetve olyan részletességű elemzések is, amiket a központi statisztikai hivatalok nem végeznek, vagy nem végezhetnek, pártatlanságuk fenntartása mellett, ezért ezeket gyakran független intézetek készítik. Ide kapcsolódik a páneurópai elemzések kérdése is, ami az Eurostat általa kezelt adatok hozzáférhetőségét igényli. A másik lényeges ok pedig az, hogy a mai kutatás-módszertan is részletes adatokat kíván, a szimulációk és modellek készítéséhez elengedhetetlen az egyéni rekordok használata. A statisztikai intézetek (nemzeti vagy szupranacionális) tehát felismerték a mutatkozó igényt, ám feltétlenül szembe kell nézniük bizonyos korlátokkal.

Az Eurostat úgy járul hozzá a kínálat bővítéséhez, hogy adatokat gyűjt a nemzeti statisztikai hivataloktól a gazdasági és társadalmi életre vonatkozóan. Ezek jól összehasonlíthatóak, harmonizált fogalomrendszerrel dolgoznak, ezért jól felhasználhatóak. Az elérésükről szóló szabályozás a nemzeti hivatalok és a Bizottság megegyezése alapján történik.

8.7.2.1.A 2002/831-es szabályozás rendelkezései az adatszolgáltatásról

Ez a szabályozás az 1997-es uniós Statisztikai törvényt veszi alapul, tárgya főként az adatok egyszerűbb hozzáférhetősége, ami még mindig a nemzeti hivatalokkal történő megegyezéshez, azok engedélyéhez kötötten jelenik meg. Mindez főként négy közösségi statisztikai termékre vonatkozik: az Európai Közösség háztartási panelfelvételére (ECHP), a munkaerő-felvételre (LFS), a közösségi innovációs felmérésre (CIS) és a Continuing Vocational Training Survey (CVTS) elnevezésű felmérésre. A szabályozás részletesen kitér ezek kutathatóságára, a hozzáférés

¹⁸⁴ Recent EU legislation for research access to confidential data – implementation and implications, Conference of European Statisticians, Geneva 10-12 June 2003

módjára. Elsősorban valamilyen kutatóintézet tagjai/alkalmazottai illetve felsőoktatási intézmények kutatói számára kívánják lehetővé tenni, más személyeknek hosszadalmas engedélyeztetési eljáráson kell átesniük a Statisztikai Bizalmassági Bizottság (Committee on Stat Confidentiality) által.

A rendelkezés szerint az Eurostat megakadályozza a nem statisztikai célú felhasználását, támogatja a tudományos célú hozzáférést, és felel az adatok biztonságáért. A Bizalmassági Bizottság ezeken kívül más mérlegelési szempontokat is bevon: a szervezet elsődleges célja, a kutatás feltételei ebben a szervezetben, a biztonsági intézkedések, illetve a kutatás eredményeinek nyilvánosságra hozási módja. Mint általában a nemzeti statisztikai hivataloknál, itt is kutatási tervet kell benyújtani a kutatás céljáról és a felhasználni kívánt adatokról, valamint a biztonsági előírások betartásáról. Ezt az adott ország (amelyhez az igényelt adatok tartoznak) statisztikai hivatalának is jóvá kell hagynia. Ezután a kutatást anonimizált adatsorokon engedélyezik, illetve az Eurostat által fenntartott kutatóhelyiségben teszik lehetővé, az outputok ellenőrzése mellett.

Ahhoz, hogy ezeket az Eurostat sikeresen alkalmazni tudja, megfogalmazzák a szükséges változtatásokat is. Így például célul tűzik ki maguk elé a megnövekedett kutatói igények felmérését, mind az adatok, mind a technikai lehetőségek tekintetében. Új megállapodásokat is létre kell hozniuk a különböző nemzeti szintű kutatóintézetekkel, és konzultálni kell a nemzeti statisztikai hivatalokkal, hiszen az Eurostat nem készít saját adatfelvételeket, csak feldolgoz, és ehhez mindenképpen meg kell teremtenie a gördülékeny együttműködés feltételeit, a szabályozás és az intézményközi kapcsolatok eszközeivel is. Fontos feladat emellett a bizalmassági kötelezettségek hangsúlyozása, előírása. Ezeket persze nem szükséges újradefiniálni, alapul vehető az egyes nemzeti hivatalok meglévő adatszolgáltatási gyakorlata (mint az Egyesült Királyságé, vagy akár az Egyesült Államoké).

Ezek az intézkedések tehát jelentősen hozzájárulhatnak (ahogyan az IMF intézkedései is) az átláthatóság növekedéséhez. Ehhez azonban szükséges a rendszer többi szereplőjének, tehát a nemzeti statisztikai hivataloknak, a

kutatóknak és még a statisztikai alanyoknak az együttműködése is. Számos új tevékenységet kíván meg például a nemzeti statisztikai hivataloktól, jóllehet a szabályalkotáskor és annak gyakorlatba ültetésekor nagy súlyt fektetnek a folyamatos konzultációra. Feltétlenül szükséges lesz azonban az outputok ellenőrzése, a biztonsági követelményeknek, a kutatási kérelmek kezelésének meghatározása az egyes hivataloknál. Néhol ez nem okoz majd gondot, máshol viszont komolyabb módszertani változtatásokat is jelenthet.

Lényeges változás az is, hogy az EU kezelésébe átlépett adatokhoz hozzáférhetnek majd nem EU-tagállambeliek is, holott némely nemzeti hivatal még EU-tagállambeli külföldiek számára sem engedélyezi azt. Mindezek a megváltozott adatkezelési körülmények felvetik a megkérdezettekhez való viszony megváltozásának kérdését is: némely országokban, a vonatkozó törvénytől függően, ez komoly jogi problémát is jelenthet.

9. A kutatási célú adatok hozzáférhetőségének hazai tapasztalatai

A kutatási célú adatokkal kapcsolatos nemzetközi gyakorlat példái megmutatták, hogy az ilyen adatok hozzáférésnek igen eltérő gyakorlata alakult ki a világ különféle országában. Ennek számos oka van, kétségtelen azonban, hogy az intézményi és jogi adottságok, illetve a személyes adatokkal kapcsolatos közhangulat meghatározó jelentőségűek. A magyarországi jogi keretek meghatározó eleméről, az Avtv.-ről a kutatás első tanulmánya részletesen beszámolt. A most következőkben azt elemezzük néhány reprezentatív példa segítségével, hogy a magyarországi adatkezelő intézmények milyen gyakorlatot folytatnak, mennyire felkészültek az adatok nyilvánossá tételére.

9.1. A választott intézmények

Az adatkezelőket két csoportra osztjuk: az adminisztratív és a statisztikai céllal felvett adatok kezelőire. A két intézménytípus gyakorlatának vizsgálatához négy konkrét esetet szerettünk volna megvizsgálni: a Központi Statisztikai Hivatal (KSH) esetében egyrészt a Munkaerőfelvételt, valamint a Népszámlálást készítő Életszínvonal és Emberierőforrás-statisztikai (ezután: KSH Munkaügy), másrészt Népszámlálási (ezután: KSH Népszámlálás) főosztályait, ezen túl a Belügyminisztérium Központi Nyilvántartó és Választási Hivatalának (BMKÖANYVH), valamint az Adó és Pénzügyi Ellenőrzési Hivatal (APEH) adatkezelését.

A KSH-n belüli két példát az alapján különítettük el, hogy hány, a Hivatalon kívül született tanulmány használja adatforrásként az említett felvételeket, feltételezve azt, hogy a tapasztalható eltérés a különféle adatközlési gyakorlat eredménye. A két különálló adatfelvételhez kapcsolódó gyakorlat mellett a KSH, mint intézmény adatkezelési stratégiáját is vizsgáltuk. Bár más választással nyilván tartalmilag eltérő eredményt kaptunk volna, a KSH hivatott olyan más, állami intézmények gyakorlatát megjeleníteni, amelyek statisztikai céllal felvett kutatási célú adatokkal rendelkeznek. Ilyen például az Állami Foglalkoztatási Szolgálat, amely a

munkanélküli regiszter adminisztratív adatai mellett kimondottan statisztikai céllal gyűjtött adatokkal is rendelkezik.

Az utóbbi két példában az APEH-et, mint ma már közzismerten fejlett számítástechnikai háttérrel sok adatot megmozgató állami hivatalt választottuk – a célnak azonban tökéletesen megfelelő választás lett volna az Országos Nyugdij Főigazgatóság (ONYF), vagy akár az Országos Egészségpénztár (OEP) is. A BMKÖANYVH-t azért tartottuk fontosnak, mert a személyi szám korlátlan használatának alkotmányellenessé nyilvánítása után ez lett az az intézmény, amely az etalon-nyilvántartásként funkcionál, és ebből a nyilvántartásból lehet a legnagyobb számú céllal lekérdezni adatokat.

Az adatgazdákat telefonon, illetve személyesen kerestük meg, változó sikerrel. A BMKÖANYVH-tól eltekintve minden adatgazda esetében eljutottunk oda, hogy legalább részben választ kapjunk kérdéseinkre. Ez utóbbi esetében, annak ellenére, hogy ez a szereplő végez leginkább piaci adatszolgáltatást (címeket értékesít direkt-marketing és piackutató cégeknek olajozottan működő vállalkozás keretében), próbálkozásaink nem jártak sikerrel. Bár az Avtv. kötelezi az adatgazdákat, hogy 15 napon belül válaszoljanak a kezelésükben levő adatokkal kapcsolatos kérdésekre, úgy éreztük, hogy ha ezzel „fenyegetőzünk”, az a kutatás céljait végeredményben kevésbé szolgálja, így letettünk róla. A BMKÖANYVH szerepét éppen ezért csak más adatokkal kapcsolatban és közvetett információ alapján tárgyaljuk.

A fenti intézmények mellett foglalkozunk egy, a kutatásunk tárgykörén kívül eső példával, a TÁRKI-val (Társadalomkutatási Informatikai Egyesülés) is. A TÁRKI figyelemre méltó képződmény: piackutató, elemző tevékenységgel foglalkozó vállalkozás, amely ugyanakkor működtet egy kutatási célú adatokat tartalmazó adatbankot is. Példája azért igen fontos, mert ugyanazok az adatvédelmi szabályok vonatkoznak rá, mint bármely más adatkezelő intézményre, viszont piaci szereplő lévén árképzésében nem kell magát a teljes ingyenességhez tartania – az nem kötelessége. Mivel ennek következtében működése egyfajta referenciaként szolgálhat, a TÁRKI példáját az összes többi előtt tárgyaljuk.

9.2. A szolgáltatott adatok köre és hozzáférése, metaadatok

Elsőként azt vizsgáljuk meg, hogy a kutatási célú adatokat kezelő intézmények milyen típusú adatokat bocsátanak az érdeklődők rendelkezésére. Feltételezzük, hogy a megszokás és egyéb tehetetlenségi tényezők mellett az adatgazdák – ismerve az adatok előállításának költségeit – annak függvényében döntenek az adatok nyilvánosságról, hogy milyen hasznosságot keletkezését várják tőle. Előzetes várakozásunk az, hogy az adminisztratív adatot kezelő szervek kevésbé hajlamosak az adatközlésre, hiszen nekik nem az adatok tudásként történő felhasználása az elsődleges tevékenységük, adatgazdává fő feladatuk ellátása mellett váltak.

Összefoglalásképpen elmondhatjuk, hogy a kutatási célú adatokat kezelő állami fenntartású intézmények konzervatív módon értelmezik az adatokkal kapcsolatos feladatukat. Minthogy mindnyájan az államigazgatás részét képezik, adatszolgáltatási palettájuk főként annak igényeihez idomul. A leggyakrabban származtatott adatok közlését tapasztaljuk, mindenekelőtt táblázatos formában, kiadványokban. A táblázat továbbfejlesztett változata a számítógépes adathordozón elérhető, zárt szerkezetű, csak a mellékelt lekérdező programmal használható adatbázis, illetve a hasonló funkciókat az Interneten prezentáló kezelő felületet. Ezek megkönnyítik az adatok használatát, a mikroadatok egyik felhasználási területét lefedik, de lényegüket tekintve továbbra is származtatott adatok maradnak.

Erős kontrasztban a nemzetközi gyakorlattal, mikroadatok szolgáltatásával egyik szervezet sem foglalkozik deklaráltan. Egyetlen kivétel a BMKÖANYVH direkt marketing és közvélemény-kutatók számára nyújtott szolgáltatása, amely azonban csak címeket közöl, így önmagában nem tekinthető elemezhető adatbázisnak. Ezzel összhangban a mikroadatok elérését segítő szolgáltatásokról, on-line elérésről vagy kutatószobáról sem beszélhetünk.

Noha egybeesik azzal, hogy a mikroadatok nem elérhetőek, nem nyilvánvaló, hogy az intézmények az egyes adatforrásokról informáló részletes metaadatokat sem szolgáltatnak. Az adatfelvételek létezésén túl csak azok az információk érhetőek el, amelyeket, ha van ilyen, az intézmény témához kapcsolódó kiadványának módszertani fejezete közöl.

9.2.1. TÁRKI

A TÁRKI tagja több igen jelentős, adatkezelőket tömörítő nemzetközi szervezetnek (amilyen például a michigani koordinációjú ICPSR), és sok nemzetközi együttműködésben a KSH-val egyenrangú intézményként jelenik meg. Az adatbank fenntartását a TÁRKI tagintézményein kívül az OTKA, az OMFB és az Open Society Institute-on keresztül a Soros Alapítvány is finanszírozza, valamint a TÁRKI Adatbank része a Nemzeti Digitális Adattárnak is, ezzel kapcsolatos forrásokat is felhasznál. Működése e tekintetben leginkább az Egyesült Államokban és Nagy-Britanniában működő kutatóintézetekhez hasonlít.

A TÁRKI Adatbank meglehetősen nagy mennyiségben tesz elérhetővé mikroadatokat. Ezek között megtalálhatók a más intézmények (például: KSH) által régebben felvett adatok, és saját felvételek is. A TÁRKI Adatbank úgy viselkedik, ahogy azt a nemzetközi tapasztalat szerint egy közérdekű adatokkal rendelkező adatgazdától elvárhatjuk.

Az Adatbank kínálatában adatokat és metaadatokat is találunk. A TÁRKI Adatbank honlapján kereső segítségével igazodhatunk el az elérhető adatbázisok között, és ha megtaláltuk, amit kerestünk, az adatok részletes leírását, valamint a felvételnél használt kérdőívet is díjmentesen letölthetjük. Meg kell jegyezni, hogy ez a leírás sokkal részletesebb, mint amit a KSH módszertani megjegyzésként tájékoztató füzetekhez mellékel. Olyan információkat tartalmaz (például a kérdések megíúsulásáról), amelyek az adatok minőségének megítélésakor igen fontosak,

amelyek alapján a kutató az adatok megszerzése, regisztráció nélkül képet alkothat arról, megfelel-e elvárásainak a kiszemelt adatbázis.

9.2.2. KSH

A KSH, mint intézmény nem foglalkozik mikroadatok szolgáltatásával, a nem államigazgatási célú adatszolgáltatás kifejezetten a származtatott adatok közlésére szorítkozik. Ezt nem csak a KSH munkatársaival folytatott beszélgetések, de KSH jelenlegi elnökével 2005 januárjában a Statisztikai Szemle számára készült interjú is világossá teszi: az elnök hosszasan taglalja a származtatott adatok elérhetővé tételéről, ugyanakkor nem szól a mikroadatokról. Mindez azonban nem jelenti azt, hogy a KSH főosztályai ne dönthetnének úgy saját hatáskörükben, hogy az általuk kezelt mikroadatokat külső felhasználók rendelkezésére bocsássák. Az ilyen lépéseket a Hivatal központi vezetése se nem támogatja, se nem tiltja.

A KSH származtatott adatainak jelentős része több úton érhető el: nyomtatott formában, elektronikus publikáció, elektronikusan lekérdezhető adat formájában és személyesen is. Bár kutatási adatokról beszélünk, amelyek értelmezése bizonyos szakmai felkészültséget feltételez, az így kialakított termékstruktúra megfelel az adatok iránt érdeklődők potenciálisan széles felkészültségbeli különbségeinek. A papíralapú információ kiszolgálja a legkonzervatívabb, esetleg számítógéppel sem rendelkező felhasználókat, míg az adatlekérdező rendszer olyan formátumban jeleníti meg az adatokat, hogy azok közvetlenül számítógépes programban legyenek használhatók. Mindezek mellett a KSH kiváló könyvtárát és szóbeli információs szolgálatot is fenntart, ahol telefonon vagy személyesen is lehet érdeklődni. Ahogy az adatközlés az utóbbi időben az ingyenesség felé tolódik el, a KSH szabad kapacitásait kihasználva igyekszik az adatszolgáltatáson túlmutató, személyre szabott értéknövelt szolgáltatásokat nyújtani: a megrendelő számára végez számításokat, készít egyedi származtatott adatokat.

A mikroadatokkal kapcsolatos politikával összhangban a KSH-ban nem létezik a mikroadatok átadására kialakított on-line környezet, vagy off-line feltételrendszer, tehát kutatószoba. Értesültünk arról, hogy a hivatal vizsgálja kutatószoba kialakításának lehetőségét, ami arra utal, hogy a mikroadatokkal kapcsolatos politikája is változhat. A megkérdezettek meglátása szerint azonban ez a folyamat régóta tart, és nem látszik gyorsulni. Ezzel együtt mégsem állíthatjuk, hogy a törekvés minden eredményt nélkülözött volna – az legfeljebb nem vált láthatóvá. Szóbeszéd útján értesültünk arról, hogy valamikor létrehoztak a Hivatalban kutatószobát, csak valamilyen, feltehetően technikai oknál fogva azt soha nem „élesítették”

9.2.2.1.KSH Munkaügy

A nemzetgazdasági szintű mutatószámok közül a KSH Munkaügy szállítja az egyik legismertebbet, a munkanélküliségi rátát. Ennek alapja az ELAR adatgyűjtési rendszer keretében végzett Munkaerőfelvétel, amely negyedévenként mintegy 60 ezer embert ér el 1992 eleje óta folyamatosan. A felvételben egy háztartás, így általában egy ember sorsa 6 negyedévig követhető, ami igen fontos, az időbeliséggel kapcsolatos vizsgálatokat tesz lehetővé.

A KSH Munkaügy a kutatók rendelkezésére bocsátja a Munkaerőfelvétel mikroadatait. A felvétel 1992-ben indult, és már 1994-től találunk olyan közgazdasági elemzéseket, amelyek a Munkaerőfelvétel alapján készültek. Ezek száma az időben folyamatosan nő, a belőle táplálkozó publikációk olyan rangos folyóiratokban jelentek meg, mint a European Economic Review. Mára elmondható, hogy a felvétel a hazai munkagazdasági kutatások egyik alappillérvé vált.

A KSH Munkaügy formális politikát nem alakított ki az adatok elérésével kapcsolatban, de szóbeli közlés alapján tudjuk, hogy minden „értelmes” kutatói igényt kielégítenek – ezt az adatokat használó kutatók megerősítik; olyanokat, akik kértek, de nem kaptak volna, nem ismerünk. Figyelemre méltó azonban, hogy a felvétel elérhetőségére a KSH honlapján, hivatalos megjelenésében nem utal

semmi, az egyfajta bennfentes tudás. Ez a politika nemcsak nem kelt fölösleges feltűnést, nem kerül szembe a Hivatal határozatlan politikájával, de a felhasználók automatikus szűrését is biztosítja.

Az adatok fizikai hozzáférése egyszerű: a mikroadatokat igénylő kutató CD- ROM-on veheti át azokat a KSH számítástechnikusaitól. On-line elérésre nincs lehetőség, ami ismét összefüggésben van a hivatali politikával, illetve a hivatalon belüli rendkívül szoros adatkezelési korlátozásokkal.

9.2.2.2.KSH Népszámlálás

A Népszámlálás a KSH lakossági felvételeinek 10 évenként felvett „zászlóshajója”: kötelező kérdéseire válaszolni és elvben minden állampolgárra kiterjed. A Népszámlálás adatbázisa fizikai szerkezetét és kezelhetőségét tekintve nem tér el lényegesen mondjuk a Munkaerőfelvétel adataitól, a fő különbséget az jelenti, hogy a lekérdezett információk száma kisebb, mint egy specializált felvételnél, a megkérdezetteké viszont óriási. Éppen ezért a Népszámlálás olyan életkori, szakmai vagy területi finomságú vizsgálatokra alkalmas, amilyenre más adatfelvétel soha nem lehet.

A Népszámlálási főosztály az utóbbi időben hatalmasat lépett előre a származtatott adatok közlése terén: az Internetről rengeteg kiadvány letölthető a 2001-es népszámlálással kapcsolatban, a részletes tájékoztatóktól, táblázatgyűjteményektől kezdve a lebonyolításhoz használt kérdőívekig. A mikroadatokat terén azonban álláspontja szigorúan megegyezni látszik a hivatal egészének álláspontjával: semmilyen formában nem kommunikálja mikroadatok hozzáférési lehetőségét, első érdeklődésre teljesen elutasítja azt. A hozzáállás eredménye látható is a tudományos publikációk terén. A 2001-es felvétel mikroadatait tudomásunk szerint kizárólag a Hivatal által készített kiadványok elkészítéséhez, illetve a főosztály munkatársainak publikációjában használták fel.

Ha figyelmünket a korábbi Népszámlálások felé irányítjuk, kiderül: azokból (például az 1990-es felvételből) láthatólag elérhetőek voltak 1-2 százalékos egyéni minták.¹⁸⁵ Ennek ellenére ezen állományok elérhetősége az érdeklődő számára ma fel sem merül. A korábbi adatkéréseket alaposabban megvizsgálva kiderül, hogy mindezzel együtt az adatok átadása egyszerűen zajlott, a hivatal munkatársainak segítőkész közreműködésével, technikailag a Munkaerőfelvétel esetében már leírt CD-ROM-os eljárással.

A Népszámlálás mikroadatainak használatára azonban nemcsak a magyarországi közgazdasági kutatás történetében találunk példát. Az Egyesült Államok-beli Minnesotai Egyetem Demográfiai Intézete által kezelt IPUMS¹⁸⁶ projekt célja népszámlálási mikroadatok összegyűjtése a világ minden pontjáról. Ebben a munkában a KSH Népszámlálási főosztálya is részt vett egy rendkívül progresszív együttműködés keretében, az 1970-es, 1980-as, 1990-es és 2001-es népszámlálás 5 százalékos mintáival. Noha ezek nem tartalmazznak földrajzi azonosításra alkalmas adatot (részleteket ld. erről a későbbiekben), a minták minden Magyarországon valaha is elérhetőnél nagyobbak: 500 ezer főről szolgáltatnak egyéni adatot. Az adatokhoz a hozzáférés ingyenes, de regisztrációt és kutatói háttérrel igényel – kapcsolatfelvétel után a felhasználók Interneten keresztül tölthetik le az adatokat.

Az IPUMS projekt igen ambíciózus: az egyes országok adatait egységes szerkezetben kívánja elérhetővé tenni, ami igen komoly munkát követel meg. A feldolgozás lassú, így bár a magyarországi adatok már a projekt birtokában vannak, azok egyelőre nem elérhetőek. Felmerül a kérdés: miért nem érhetőek el ezek az adatok a Népszámlálás magyarországi honlapjáról? Feltételezésünk szerint azért, amiért egyetlen mikroadat elérhetőségéről sincs információ – a feltételezett okokat később részletezzük. Mindenesetre meglepő az, hogy egy kulcsra kész adatforrás elérhetősége kérdéses.

¹⁸⁵ Lásd például a mikroadatok nyilvánosságának hasznait tárgyaló fejezetben is hivatkozott Kertesi (2000)-t.

¹⁸⁶ Integrated Public Use Microdata Series, Integrált Szabadon Felhasználható Adatsorok

Végül említést kell tenni egy fontos információról, még ha hipotetikus helyzetre vonatkozik is. Korábban már említettük, hogy a KSH-ban létrejött, de soha nem működött kutatószoba, illetve hogy a Hivatal munkatársai dolgoznak kutatószoba felállításának lehetőségén. Megkérdeztük a Népszámlálási Főosztály munkatársait, látnak-e arra lehetőséget, hogy ha ilyen, a nemzetközi gyakorlatban bevett kutatószoba létesül, ott az érdeklődő kutató megfelelő jogi nyilatkozat aláírása után a Népszámlálás teljes anyagával. Noha egy ilyen konstrukció merőben más adathozzáférésre teremt lehetőséget, mint a ma megszokott, a válasz pozitív volt: a lehetőség ellen a Főosztálynak nincs kifogása, de arról a Hivatal vezetőségének kell döntést hoznia.

9.2.3.APEH

Az APEH olyan kormányzati szerv, amelynek nem feladata az általa kezelt adatok nem-adminisztratív felhasználása. Ezt az önazonosságot tükrözi az intézmény (egyébként ügyintézésre igen jól használható) honlapja, és az ott dolgozók attitűdje is. Az APEH birtokában ugyanakkor rendkívüli információtömeg van: kezeli minden bevallást készítő adózó minden bevallási időszakra vonatkozó személyi jövedelemadó, vállalkozási nyereségadó- és EVA-adatát, hogy csak a legfontosabb példákat említsük. Ezek az adatok önmagukban is igen értékesek kutatási célokra, az adó azonosítók által és a BMKÖANYVH azonosító-kulcsain keresztül lehetséges adat-összekapcsolás azonban megsokszorozza értéküket (lásd később).

Formális úton mikroadatok elérhetőségéről nem lehet információt szerezni, ilyen lehetőségre az APEH honlapja nem utal. Csak személyes beszélgetés során, és csak többszöri rákérdezés után derül ki, hogy adatkérések voltak a múltban és lehetségesek is, ha arra megfelelően megalapozott igény van. Ezeket a Hivatal csak akkor szolgáltatja, ha a feldolgozáshoz szükséges gépidő rendelkezésre áll, és természetesen anonimizálva.

9.2.4. BMKÖANYVH

A BM Nyilvántartó kutatási célú adatokat közvetlenül nem kínál, csak olyanokat, amelyek egy kutatás cél-népességét segítenek kiválasztani (mintát venni). Az erről szóló minden információ megtalálható a BMKÖANYVH honlapján.

9.3. Adatkérők kormányzaton belül és kívül

A kutatási célú adatokat mind a kormányzat, mind a piac, mind a kutató szféra használja, a megvizsgált szervek mindegyike kiszolgálja ezt a három szegmenst. Az adatigények erősen eltérőek: a kormányzati szervek majdnem minden esetben származtatott adatokat igényelnek, a kutatók főként elemi adatokat, a piaci szereplők pedig mindkettőt (bár elemi adatok iránti igényük főként a település-szintű adatokra korlátozódik). Noha a lehetősége megvan rá, az információcserében betöltött szerepe miatt ennél az adattípusnál ritkának tűnik a lakossági érdeklődés – a számunkra legfontosabb mikroadatok esetében ilyenről nem is beszélhetünk.

Amint azt már említettük, a felhasználókkal kapcsolatban több ízben felmerült a „kontár probléma”, az a kétely, hogy vajon képesek-e megfelelően értelmezni, használni a nekik kiadott adatokat. Emellett az adatgazdák több esetben beszámoltak az adatok visszaszolgáltatásáról. Főleg a piaci, nem kutatói szándékú felhasználók esetében fordult elő, hogy az adatgazdától vásárolt elemi adatokat olyan formában használták fel, ami szakmailag erősen kifogásolható. Ilyen esetekre az adatgazdák érezhetően nincsenek felkészülve, biztonságosabbnak érzik a kontárok kiszűrését már az adatkérés folyamán.

9.4. Árazási elvek, üzletpolitika, adatkezelési költségek

Az adatkezelők árazási politikája a rendelkezésre bocsátott adatok típusával szoros összefüggésben alakul. Minthogy legtöbbjük formálisan csak származtatott adatot szolgáltat, jellemzően csak az ilyen adatnak van szabott ára. Több esetben találkozunk ingyenes adatszolgáltatással is.

9.4.1. TÁRKI

Mint azt leírtuk, a TÁRKI árazását nem köti semmiféle jogszabály, saját belátása szerint árazhatja az adatait. Az Adatbank árai világosak: az adat és a felhasználó jellegétől is függnék. Minden adatbázisnak van egy alapára, ami aszerint eltérő, hogy új (egy éven belüli) vagy régi felvételtől (4 évnél régebbi) van szó, illetve hogy az egyszerű, vagy bonyolult (több vagy kevesebb ismerv található meg benne). Természetesen a jobb minőségű – frissebb, részletesebb – adatbázisért állapítanak meg nagyobb alaparat, ami nettó 30 és 300 ezer forint között mozog. Ebből három fokozatban kaphat a felhasználó kedvezményt: külföldi kutató, diák az alapár 60 százalékát, hazai, nem TÁRKI-tag kutató vagy diák annak 30 százalékát köteles fizetni, míg tagintézmények kutatójának vagy diákjának a hozzáférés ingyenes.

A TÁRKI árazási politikájának két jellemzője emelendő ki: egyrészt nem határkötség alapú, másrészt differenciál. Az alkalmazott differenciálás már eleve kizárja a határkötség alapú árazást, hiszen az adathozzáférés minden elkülönített esetben ugyanúgy történik: tökéletesen előkészített adatot vesznek le a „polcra”, majd tesznek elérhetővé az érdeklődő számára. Az árképzés a korábban kifejtett kutatás által generált externália-logika kiterjesztett változatán alapuló diszkrimináció és a kétrészes árazás keveréke. Az előbbi szerint nem csak a kutatás preferálandó a nem kutatási célú felhasználással szemben, de feltehetjük, hogy a hazai felhasználásnak több hazai externális hatása van, mint a külföldinek – az olyan kutatásoknál, amelyek az országra jellemző sajátságokra koncentrálnak, ez igaz is lehet. Ennek értelmében racionális a kevésbé közjószág jellegű terméket előállító felhasználótól nagyobb árat kérni. Egészen az árskála aljáig nincs fix árkomponens, ott viszont a hozzáférés határkötsége nulla: a TÁRKI tagok tagdíjukért cserébe nulla egyedi költséggel, tehát határkötségen jutathatnak hozzá az adatokhoz. Ez az árazási stratégia tökéletesen racionális egy kemény költségvetési korláttal működő intézmény számára.

A differenciálásnak azonban van egy másik, az alapár kialakításáért felelős dimenziója is: az Adatbank minden adata korlátozás nélkül hozzáférhető, de az újabb adatok drágábbak. Elvileg az egy éven belül felvett adatok ára a legnagyobb, 2005 decemberében azonban az Adatbankban még nem található 2005-ös adat.

Ennek az árazásnak a megítélésekor figyelembe kell venni azt az intézményei kontextust, amiben a TÁRKI Adatbank működik, nevezetesen a TÁRKI-t, mint kutatóintézetet és egyesületet. A TÁRKI Adatbankba számos adat a Kutatóintézet üzletága által felvett adatfelvételtől kerül be. Ezek célja rendszerint olyan tanulmány készítésének támogatása, amelyet a TÁRKI megrendelésre készít, vagy amellyel tudományos reputációját igyekszik növelni. Nyilvánvaló, hogy ha a házon belüli kutatók még nem sok publikációt jelentettek meg, védeni kívánják a saját kutatási céljaikra felvett vagy vásárolt adatokat azáltal, hogy magas árat szabnak. Magatartásuk racionális: mivel fő tevékenységük egyike az adatok értéknövelt hasznosítása, tisztos részt kérnek abból, ha valaki hasonlóval próbálkozva vetélytársuk lesz. Ezen a ponton két megjegyzést kell tennünk. Egyrészt az „első vonalat” nem csak a házon belüli felhasználásra, hanem minden TÁRKI tagra kiterjesztik – akármilyen magas is legyen az alapár, ezek az adatkérők 100 százalékos kedvezményben részesülnek. Másrészt igaz, hogy magasabb árat szabnak a potenciális vetélytársaknak, de csak igen rövid távon zárják ki a verseny lehetőségét, zárják el az adatokat teljes mértékben a hozzáféréstől.

A TÁRKI Adatbank költségstruktúrájáról nincs közvetlen információnk. Mivel azonban az adatbank elkülönült egység, amely adatfelvételt nem végez, költségei meglehetősen stabilak. Az Adatbanknak deklarált feladata a rá bízott adatok karbantartása és hozzáféréseinek biztosítása, ami ha az adatállomány nem indul robbanásszerű növekedésnek, egyszerű és állandó költségstruktúrát eredményez. Az adatbázisnak a kezdetektől feladata a felhasználók hozzáféréseinek bonyolítása, a folyamatos proaktív munka következtében pedig az egyes tranzakciók határköltsége nulla.

9.4.2.KSH

A KSH adatközlési rendszere jelentős átalakuláson ment keresztül az utóbbi években. Az Avtv. rendelkezéseit értelmezve a KSH vezetői arra a következtetésre jutottak, hogy a leghelyesebb, ha a Hivatal kezelésében levő származtatott adatokat nyilvánossá teszik. Ma már számos, korábban pénzért vásárolható kiadvány tölthető le a KSH Internetes honlapjáról, nem egy esetben oly módon, hogy az eredeti papír alapú anyag már nem elérhető. Szintén az Internetről szerezhetők be olyan adatok, táblázatok, amelyeket a felhasználó saját igénye szerint állíthat össze, majd tölthet le adatkezelő programmal való használat céljára. Emellett elérhetők származtatott adatok kiadvány formájában, amelyek áráról nyomdaipari tudás hiányában nem tudjuk megállapítani, hogy az előállítás határköltségét testesítik-e meg, de ez nem kizárt, mert nem különösebben drágák.

A KSH tehát végső információforrásként pozicionálja magát a piacon. Nem a kiemelkedő adatgyűjtő, hanem végső adatforrás szerepét hangsúlyozza. Szolgáltatásainak ára az Avtv.-vel összhangban (helyesen) diszkriminatív: az alapvető származtatott adatokat ingyen adja (a személyes és telefonos szolgáltatás egyedülálló módon még az Internet elérésének fix költségét is megtakarítja), az „értéknövelt” szolgáltatásokért viszont pénzt kér.

Árazási kérdésekről lévén szó, igen fontos explicitté tenni, hogy végeredményben a kívülág hozzáféréseinek korlátozása is egy szélsőséges kétrészes árazás, amelyben a hozzáférés határköltsége nulla, a fix költsége pedig az, hogy az érdeklődőnek a KSH-ban kell dolgozni. Ismereteink szerint nem sok olyan tudományos kutató van, aki hajlandó ezt a fix költséget megfizetni.

Ha fenntartjuk elképzelésünket, hogy az adatgazda a nyilvánosságra hozásról az adatok előállításának ismert költségei és várt hasznai alapján dönt, két indirekt következtetésre juthatunk. Ha az externális hatások nyomán indulunk el, azt gondolhatjuk, hogy a Hivatal főszabály szerint azt gondolja: a mikroadatok hozzáférhetőségéből származó társadalmi hasznok általában alulmúlják azt a

költséget, amit a kontárok kiszűrése és a szükséges adatvédelmi intézkedések, illetve közönségkapcsolati adminisztráció felállítása jelentenek. Az adatok előállításának költségét egy ilyen kalkulációnál természetesen nem kell kiadásként figyelembe venni, hiszen az már elsüllyedt költség. A nagy költség sokkal inkább motivációként kellene, hogy szolgáljon olyan vállalkozások beindítására, amelyek pozitív nettó eredménnyel kecsegtetnek.

Egy másik lehetőséget a TÁRKI árazási modellje sejtet: elképzelhető, hogy az árazást az az – állami intézmények esetében elfogadhatatlan – cél motiválja, hogy védettséget biztosítson az adatokon keresztül saját munkatársainak a külső felhasználók versenyétől a kutatási eredmények piacán. Bár nem tudjuk, hogy a két motiváció közül melyik a valódi, vitathatatlan, hogy a költségekről a nemzetközi tapasztalatok alapján kialakított képpel, illetve az egyes adatok felhasználásával készült tanulmányok hazai piacán kialakult versennyel a második megoldást valószínűsítik.

9.4.2.1.KSH Népszámlálás

A Népszámlálási Főosztály politikája formálisan megegyezik a KSH egészével: az adatfelhasználók felé elsősorban származtatott adatokon keresztül jelenik meg, amelyek jelentős része korlátozás nélkül hozzáférhető az Interneten, ami megfelelő költségvetés esetén optimális megoldásnak tekinthető. A főosztály emellett megrendelésre is készít adatfeldolgozásokat, amelyek árazásában elsősorban a befektetett munkaidőt érvényesítik.

A Népszámlálási főosztálynak a korábban ismertetett helyzetnek megfelelően nincs explicit árazási politikája a mikroadatokra vonatkozóan. Árazásra két példán keresztül következtethetünk, amelyekből az egyik egy historikus példa¹⁸⁷, a másik pedig a már említett, a Minnesotai Egyetemmel kötött megállapodás. Ismereteink szerint a múltban lezajlott adathozzáférés, az adatbázis átadásának költsége nem

¹⁸⁷ Az adatvédelmi biztos 170/K/1998-as állásfoglalása az MTA Közgazdaságtudományi Kutatóintézetének a Mikrocenzus iránti igényét az adatok megfelelő kezelése mellett támogatja. Ennek nyomán valósult meg az egyik adatátadás.

volt igazán jelentős, akkori árakon mintegy 100 ezer forintról beszélhetünk. Tekintve a minimális adatelőkészítési igényt, ez az összeg kis túlzással tekinthető az adathozzáférés határkötségének.

A Minnesotai Egyetemmel kialakított együttműködés igen jó példáját adja annak, hogyan érdemes gazdálkodni az adatvagyonnal. Szakmailag értékes együttműködésről van szó, amely nemcsak a Hivatal reputációját növeli, de anyagi bevételt is jelentett. A pontos hivatkozás kedvéért szeretnénk volna betekinteni az adatok átadását szabályozó szerződésbe, mert az részletesen szabályozza az adatok átadását, meghatározza a közölt információ tartalmát, az anonimizálás módját. Az alkalmazott technika ismerete azért lenne fontos, mert könnyen hozzáférhető alapul szolgálhatna későbbi adatkérések teljesítése esetén. Az adatokat a KSH értesülésünk szerint nem díjmentesen adta át. Az átadás díjának ismerete segítséget nyújtott volna atekintetben, hogy megismerjük a Népszámlálás árazási stratégiáját. A vonatkozó szerződés kiadását a Hivatal jogi osztálya mérlegeli, és döntésre a tanulmány lezárásáig nem jutott.

Az együttműködés az adatok egyszeri hatékony értékesítésén túl felvet kérdéseket is. Korábban említettük, hogy az adatok külső hozzáféréseinek korlátozása két módon értelmezhető, melyek közül az egyik az, hogy az adatkezelő a közzétételtől negatív hozadékat vár. Ennek oka lehet a várakozások szerint kis társadalmi hasznosság, vagy a közvetlenül a közzététellel összefüggő költségek túlzott nagysága. A Népszámlálás 2001-es mintája esetében a hozzáférés költségeiről két információnk van: egyrészt a korábbi minták kiadásának költségei, másrészt az, hogy a Minnesotai Egyetem számára már elkészítették a felvétel adatvédelmi szempontból megfelelő változatát. Mivel egy újabb, változatlan tartalmú adatkérés esetében egyik költség sem jelentős, csak arra következtethetünk, hogy a kedvezőtlen döntés oka a kicsinek becsült társadalmi haszon.

Korábban említettük: bár a mikroadatok hozzáféréseinek szolgáltató-oldali költségét nem befolyásolja az a tény, hogy egy államilag finanszírozott adatfelvétel drága, minden mástól eltekintve valószínűbbé kellene, hogy tegye annak nyilvánosságát. A

Népszámlálást a legdrágább hazai felvételnek neveztük, bár valójában nincsen pontos információnk arról, hogy ténylegesen mennyibe kerül az előállítás. Ennek egyik oka, hogy bármilyen adatfelvétel költségét igen nehéz összegyűjteni, hiszen a Hivatal hatékonyságát az biztosítja, hogy egyes feladatokat specializált részlegek intézik. A másik ok az, hogy a ma már működő egységes könyvelési rendszer 2001-ben, illetve előtte még nem üzemelt. Annyit azonban tudunk, hogy a 2005-ös Mikrocenzus 560 millió forintos költséggel 190 ezer emberről gyűjtött adatot (lásd: magyarorszag.hu), tehát az egy főre jutó adatgyűjtési költség mintegy 2900 forint volt. Ez a szám természetesen nem szorozható be automatikusan 10 millióval, hiszen a Mikrocenzus a Népszámlálásnál jóval több kérdést tartalmaz és a szervezésben méretgazdaságosság is érvényesül. Tételezzük fel, hogy e két tényező összességében megfelel a költségeket. Így a költség mintegy 15 milliárd forintra becsülhető.

A pontos számoktól függetlenül Népszámlálás és a Mikrocenzus adatfelvétele tehát igen költséges vállalkozás. A múltbeli tapasztalatok a Minnesotai Egyetemmel létrejött együttműködés és a TÁRKI példája alapján bátran kijelenthetjük: az adatok nyilvánosságának költsége elenyésző az adatok létrehozásának költsége mellett, hiszen az semmilyen összefüggésben nem áll az adatbázis nagyságával. Ennek értelmében igen erőteljes az a nyomás, ami az adatok nyilvánosságra hozatala mellett a döntéshozókra nehezedik.

9.4.2.2.KSH Munkaügy

A KSH Munkaügy árképzése igen hasonlít ahhoz, amit a Népszámlálás múltbeli adatközlései esetében tapasztaltunk, azzal a lényeges különbséggel, hogy folyamatosan üzemel. Itt az adatok hozzáféréseinek nincs formális akadály, és ismereteink szerint az adatok létezésének informális terjedése kellően hatékony szűrőt biztosít ahhoz, hogy az adatokat kontároknak vagy rossz szándékú felhasználóknak eszébe jussok elkérni.

Az adathozzáférésnek nincs szabott díja, az árképzésben két elv érvényesül. Egyrészt megjelenik az externália-gondolat: a külföldi felhasználók rendszerint nagyobb áron juthatnak a mikroadatokhoz, mint a hazaiak, mivel utóbbi felhasználásoktól nagyobb társadalmi hasznokat remélünk. Másrészt alacsonyabb (nulla) árat szabnak a kutatóknak, akik észrevételeikkel hozzájárulnak a felvétel jobbá tételéhez, amely a KSH-ra visszaható externális hatás.

9.4.3.APEH

Az APEH árazása gyakorlatilag a KSH Munkaügy árazását tükrözi. Egyfelől az adathozzáférés lehetősége nem publikus, ami szűrőként funkcionál, másrészt az árban csak az adathozzáférés költségét érvényesítik.¹⁸⁸ Ugyanakkor az APEH adataira csak igen kevés tudományos igényű tanulmány épít.

9.4.4.BMKÖANYVH

A BM Nyilvántartó árazása egyértelmű, a honlapján megtalálható tájékoztatóból pontosan leolvasható. A címnyilvántartás 470 forintért szolgáltat egy nyilvántartási adatot, de 5, vagy kevesebb adat esetében az 5 főért járó díjat kell fizetni. Diszkrimináció semmilyen formában nem érvényesül, bár a kormányzati szervek ingyenesen férnek hozzá a Nyilvántartó adataihoz.

9.5. Adatvédelem

Az adatvédelem terén a magyarországi Avtv. nemzetközi összehasonlításban igen markáns véleményt képvisel, amikor úgy rendelkezik, hogy egy adat mindaddig személyesnek tekinthető, ameddig bármekkora költséggel kapcsolatba hozható azzal a személlyel, akire vonatkozik. A nemzetközi gyakorlatban láttuk, hogy a feltétel általában ennél szelídebb, és csak azt követeli meg, hogy az egyének ne

¹⁸⁸ Ez az információ szóbeli közlésből származik, a pontos kimutatás a tanulmány lezárásáig nem érkezett meg a Hivataltól.

legyenek belátható energiával azonosíthatók.¹⁸⁹ Ennek megfelelően ma Magyarországon számtalan olyan adatfelvétel létezik, amely alapján rendőrségi, titkosszolgálati és hasonló eszközökkel, rendkívüli erőbedobással beazonosítható az adat alanya, egy kutató számára azonban ez többnyire megoldhatatlan feladatot jelent.

A magyarországi gyakorlatot a KSH esetében az Avtv. és a Statisztikai törvény, a többi adatgazda esetében csak az előbbi szabályozza. Az adatátadási gyakorlat tiszteletben próbálja tartani a törvényt, de látni kell, hogy mikroadatok esetében erre a legteljesebb mértékben nincs lehetőség. A magyar gyakorlat sokkal inkább az ésszerű nemzetközi gyakorlattal, mint az Avtv. feltételeivel van összhangban: az egyén hivatalos azonosításra alkalmas adatait (személyi igazolvány szám, TAJ, adószám, stb.) egyetlen statisztikai célú felvétel sem rögzíti, az adminisztratív adatforrásokról pedig minden statisztikai célú felhasználás esetben törölni kell azokat. Az azonosíthatóság teljes körű ellenőrzését általában nem végzik el, de a józan ésszel és szakmai szemmel relevánsnak tűnő dimenziókban minden általunk ismert esetben olyan mértékben csökkentik az információk mértékét, hogy az azonosítás költsége az adatokhoz hozzáférő emberek elvárható viselkedését feltételezve prohibítív legyen.

De mit is jelentene mikroadatok esetében az Avtv.-nek való teljes megfelelés? Azt, hogy az adat gazdája a megfigyelhető változók minden lehetséges kombinációját előre létrehozza, és ellenőrzi, hogy nincs-e – a Statisztikai törvény szabta határt használva – 3 megfigyelésnél kevesebb az adott cellában. Ez természetesen lehetetlen feladat, így erre egyetlen olyan adatgazda sem vállalkozik, aki átad mikroadatokat. Ezzel együtt minden esetben tapasztaltunk törekvést arra nézve, hogy az egyedi azonosításban kulcsszerepet betöltő finom területi azonosítók ne szerepeljenek az adatbázisban.

¹⁸⁹ Az azonosítás lehetőségére mi sem mutat rá pontosabban, mint az a nyilatkozat, amit a Health and Retirement Study használóinak kell aláírniuk (megtalálható a http://ssl.isr.umich.edu/hrs/reg_cou.html címen). Ez az első pontban kötelezi a nyilvános (!) minta használóját, hogy nem kísérli meg a kérdezettek azonosítását: „Make no attempts to identify study participants”.

Egyetlen konkrét példát ismerünk arra nézve, hogy az Avtv. előírásait pontosan be akarják tartani, ez pedig a KSH Interneten publikált, elsősorban településszintű adatai. Beszélgetéseink során azt az információt kaptuk, hogy törekszenek az adatvédelmi aggályok teljes kiküszöbölésére, de az igen nagy energiát igényel.

A kialakult rendszer működése mellett szól, hogy nincs olyan visszaélésről tudomásunk, amely az egyedi adatok beazonosításán alapult volna (míg a formálisan meghirdetett adathozzáférési lehetőségek esetében egyéb visszasságokról, főleg az üzleti felhasználók esetében, tudunk). Vajon mi lehet ennek az oka? Véleményünk szerint az, hogy a mikroadatok hozzáférésénél erőteljes szakmai szűrő működik, ami eleve olyan felhasználókat enged csak az adatokhoz hozzáférni, akiknek eszükbe sem jut nem megfelelően használni az adatokat. A lelepleződés költségei ugyanis igen magasak, mivel a kutató ebben az esetben a jövőre nézve szakmailag ellehetetleníti saját magát.

Mindez azonban fontos üzenet azoknak az intézményeknek, amelyek még nem adnak át mikroadatot: valamilyen módon védeniük kell magukat a kétes szándékú felhasználóktól. Hasonló következtetés vonható le arra az esetre, ha a vizsgált intézmények a mikroadatok hozzáférhetőségét nyíltan zászlajukra tűzik. Az adatkérők számának növekedésével óhatatlanul megnő a kétes felhasználás valószínűsége, ami ellen jogi eszközökkel, illetve a szűrés formalizálásával érdemes védekezni. Bizonyos esetekben megfontolandó az adatok torzítása, a jelenleg nem elérhető információk megnyitása esetében pedig a kutatószoba kialakítása. Ez természetesen növelni fogja az adatvédelem ma nullához közeli költségét, de a legtöbb esetben koránt sem annyira, hogy értelmetlenné tegye az adatok megnyitását.

9.5.1. TÁRKI

A TÁRKI minden felhasználójával nyilatkozatot írat alá az adatok átadása előtt. Ez a nyilatkozat tudományos felhasználás esetén rögzíti, hogy az adatokat csak ilyen

célra lehet használni, de nem tartalmaz adatvédelmi klauzulát. Minden egyéb felhasználási cél esetében egyedi nyilatkozat aláírására szükséges. Az adatok effektív védelmét a minta területi szórtsága, az információk korlátozottsága, és az egyéni azonosítók hiánya biztosítja.

9.5.2.KSH

A KSH legújabb fejlesztése a KSH származtatott adatainak on-line elérhetősége. Ez a rendszer az adatok érzékenységet adatvédelmi szempontból nem változtatta meg, csak korábban egymással kapcsolatban nem levő adatokat azonos felületen tett elérhetővé, így az esetleges problémák szembeötlőbbek. Mint azt korábban már írtuk, az adatvédelmi aggályok tökéletes kiküszöbölése rendkívüli feladat, de a Hivatal kísérletet tesz rá, vállalva az ezzel járó igen jelentős költségeket.

9.5.2.1.KSH Népszámlálás

A Népszámlálási Főosztály azt, hogy nem ad ki mikroadatokat, mindenekelőtt adatvédelmi aggályokkal indokolja, formális elveket így ez ügyben nem alakított ki. Az árazáshoz hasonlóan, itt is csak a múltbeli tapasztalatokra hagyatkozhatunk.

A hazai adatátadások esetében értesülésünk szerint a Főosztály a standard gyakorlatot követte. Az adatok effektív védelmét a minta területi szórtsága, az információk korlátozottsága, és az egyéni azonosítók hiánya biztosítja, természetesen a hozzáférésre jogosultak körének erőteljes szűrése mellett.

A Minnesotai Egyetemmel kötött együttműködés adatvédelmi szempontból igen érdekes példát szolgáltat. Mivel nem a Népszámlálási Főosztály kontrollálja az adathoz hozzáférők körét, a szokásosnál alaposabb vizsgálatnak vetette alá a felvételt. Egyfelől törölték az összes földrajzi azonosítót, ami egy 500 ezer fős mintánál akár túlzott óvatosságnak is tekinthető: egy kistérségre átlagosan mintegy 3 ezer fő jut, azonosításuknak esélye nulla. Másodlagos intézkedésként nem teljes

körűen, de végigellenőrizték mindazokat a változókat, amelyek kategóriájába kevés megfigyelés jut, és összevonták azokat.

A módszer ismeretében azt mondhatjuk, hogy a mintavétellel keletkezett adatbázisok esetében ez a megoldás megjelöli az adatvédelmi intézkedések ésszerű maximumát. Egyúttal példát is szolgáltat: az eljárás kivitelezhető, nem beszélhetünk arról, hogy költsége prohibitív lenne. Ennek jelentősége annál nagyobb, mert a mikroadatok kiadása elleni legfontosabb érv az, hogy az adatvédelmi költségek igen jelentősek.

9.5.2.2.KSH Munkaügy

A Főosztály ugyanazt a gyakorlatot követi, mint amit a Népszámlálás esetében láttunk.

9.6. Adatok összekapcsolásának lehetőségei: a BMKÖANYVH és egy példa

Az adatbázisok összekapcsolás rendkívüli elemzési lehetőségeket rejt magában. Példaként tegyük fel, hogy az öregedés és a munkavégzés közötti kapcsolatot szeretnénk megvizsgálni. Ehhez az 55 és 70 év közötti emberek munkavállalását szeretnénk vizsgálni. Ez a kutatási helyzet számos problémát vet fel:

1. az általános célú felvételekben nincs megfelelő számú eset arra, hogy egy ennyire szűk korcsoport viselkedését vizsgáljuk,
2. a nyugdíjkorhatár körüli viselkedésre komoly hatással van az, hogy egy ember korábban milyen módon állt helyt a munka terén, ugyanakkor az általános felvételekben nincs olyan retrospektív adat, amely kellő pontossággal leírná az egyének munkatörténetét, végül
3. az adott korcsoport viselkedését alapjaiban határozza meg egészségi állapota.

Ezen problémák önmagukban is olyan adatigényűek, hogy igen nehéz, vagy lehetetlen megfelelő adatbázist találni hozzá. Együttes elemzésük azonban vagy

igen nagy költséggel megvalósítható specializált adatfelvételt, vagy adminisztratív adatok használatát igényli.

Fontos tisztázni, hogy statisztikai céllal gyűjtött adatok összekapcsolására nincs lehetőség, mert az adatfelvételeknél erre alkalmas azonosítókat nem gyűjtenek. A megfelelő adminisztratív adatok összekapcsolásával azonban a tényezők önálló, és egymáson keresztül hatása is kimutatható, így ismét olyan eredmények születhetnek, amit más adatok segítségével nem.

Tudomásunk szerint egyetlen olyan adatkérés történt Magyarországon, ami adminisztratív adatbázisok összekapcsolását igényli. A Pénzügyminisztérium Gazdaságpolitikai Főosztályának Közgazdasági Kutató Osztálya fordult a BMKÖANYVH-hoz 2005 januárjában azzal a kéréssel, hogy működjön közre adminisztratív adatok összekapcsolásában. A kérést az állami gazdálkodás hatékonyabbá tételére irányuló kutatás támogatásában megjelölve az alábbi eljárási javaslatot tette a BMKÖANYVH -nak (itt: BM Nyilvántartó, BMNy):

Az „... adatbázis létrehozásához szükséges az Országos Egészségpénztár (OEP) és az Adóhivatal (APEH) adatállományaiban található egyéni adatok összekapcsolása, a kiválasztott mintára kiterjedő körben. Az összekapcsolás közvetlenül nem lehetséges, mivel a két hivatal egymás adatait kutatási célra személyes azonosításra alkalmas módon nem kapcsolhatja össze, illetve az összekapcsoláshoz szükséges azonosítókkal nem rendelkezik.

Az összekapcsolás ezért az alábbi módon történhet meg:

1. Az OEP leválogatja a 40 ezres mintát, és minden elemének ad egy sorszámot.
2. Az OEP az egyénekhez tartozó TAJ számot helyettesítő kapcsolati kódot és a megfelelő sorszámot átadja a BM Nyilvántartónak (BMNy).

3. A BMNy kikeresi az ezekhez a kapcsolati kódokhoz tartozó adószámot helyettesítő kapcsolati kódot, és ezt a kódot a megfelelő sorszámmal együtt eljuttatja az APEH-nek.
4. Az APEH a kapcsolati kód alapján leválogatja a megfelelő egyének jövedelmi és befizetési adatait, törli a mintából a kapcsolati kódot, de megtartja a sorszámot, és így adja át az adatállományt a PM-nek.
5. Az OEP szintén törli az eredeti mintából a kapcsolati kódot, de megtartja a sorszámot, és így adja át az adatállományt a PM-nek.
6. A PM az OEP által generált sorszám alapján a két hivataltól kapott adatállományt összekapcsolja.”

Megfigyelhető, hogy az adatkérő képes volt olyan technológiát konstruálni, amely egyszerre felel meg az Avtv.-nek, és szolgáltat a céljainak megfelelő adatokat. Ez azért figyelemre méltó, mert az adminisztratív adatok kezelői a mikroadatokra vonatkozó adatkérések esetében minden más érv előtt az adatvédelmi nehézségekre szoktak hivatkozni. A fentihez hasonló kérések teljesüléséhez azonban nem elégséges az adatvédelmi szempontból kielégítő technológia: az adatkérést a BMKÖANYVH jogszabályi megokoltság hiányában visszautasította.

10. Összefoglaló gondolatok

Közgazdasági elemzésünk fő eredménye az, hogy ha finanszírozható, akkor társadalmilag optimális a közérdekű adatokat nyilvánosan hozzáférhetővé tenni. Ennek értelmében az adat árának mindössze az információátadás igen kicsi határköltségét kell tartalmaznia, amely javaslat egybeesik az Avtv. előírásával is. Írásunkban ugyanakkor számos példát mutattunk arra, hogy az adatkezelő intézmények árazási gyakorlata sokban eltér ettől a forgatókönyvtől, bár a rendelkezésre álló adatokból nem tudtuk eldönteni, hogy egyszerű monopolista árazásról, vagy a társadalmi jólétet maximalizáló, effektív költségvetési korlát esetén optimális magatartásról van szó. Ebben a záró fejezetben először azt foglaljuk össze, hogy milyen következményekkel járna, ha a jogi szabályozás nem változna, de az adatszolgáltatás közel ingyenessé tétele hirtelen minden intézményre vonatkozóan kikényszeríthető lenne. Megvizsgáljuk azt is, hogy ez az ingyenesség mennyiben valósítja meg az Avtv. egyik fő célját, az információszabadságot. A fejezet második részében pedig összefoglaljuk ajánlásainkat, amelyben már a jogi környezetet is alakíthatónak tételezzük fel.

10.1. Gondolatkísérlet: a kikényszerített teljeskörű ingyenes adathozzáférés következményei

10.1.1. A keresleti oldal

A klasszikus adatok esetén nagy hangsúlyt fektettünk annak bemutatására, hogy a hozzáférésnek nem csak pénzben mérhető költségei (ára) van, hanem igen jelentősek a nem pénzbeli változó költségek (egy jellemző dimenzió volt az adatkérések időigénye), illetve a hozzáférés technológiájának (például Internet-hozzáférés, értelmezésre és adatkeresésre vonatkozó szakértelem) fix költségei. Csalóka tehát azt állítani, hogy az adathozzáférés árának nullára csökkentésével megvalósul az adatok teljes nyilvánossága, amíg a további költségek még mindig komoly belépési korlátokat állítanak egyes adatkérések elé.

Kétségtelen, hogy a pénzben kifejezhető változó költségek csökkenése egyes eddig távolmaradó személyeket arra ösztönözhet, hogy most már befektessenek az adatigénylés technológiájába, mivel nettó hasznaik immár nagyobbak a beruházás fix költségénél. Ez azonban korántsem jelenti azt, hogy minden állampolgár számára megéri igénybe venni ezeket az adatigénylési eszközöket, vagyis a közérdekű adatok továbbra sem bizonyulnak klasszikus közjóságnak. Az adathozzáférés pénzben kifejezhető ingyenességének fő kedvezményezettjei azok a szereplők, akik már kifizették az adatkérési technológia fix költségeit, az ő esetükben egyértelműen számolhatunk azzal, hogy az ő adatigényléseik száma megnövekszik.

Ennek az eredménynek a szem előtt tartása igen fontos annak tükrében, hogy az adatok nyilvánosságra hozatalának különösen propagált módja azok Internetes publikálása. Ez a módszer véleményünk szerint két okra vezethető vissza. Egyrészt az internetes publikálás az információszabadság kérdésében példaadó országok (például skandináv országok, Egyesült Államok) gyakorlatában is a leggyakrabban használt módszer. Másrészt az ilyen módon való nyilvánosságra hozatal az adatkezelő költségei szempontjából sokkal hatékonyabb, mint az egyéb módszerek. Az első motivációt ugyanakkor Magyarországon óvatosabban érdemes alkalmazni, mivel hazánkban az Internet keresőtechnológiaként való alkalmazása sokkal kevésbé elterjedt, mint a példaként állított országokban.¹⁹⁰

A fix költségek csökkentésére számos lehetőség kínálkozik, de a belépési korlátok teljesen nem tüntethetők el. Klasszikus adatok esetében ilyen magától értetődő módszer a nyilvános Internet-hozzáférés lehetőségeinek növelése, ami ugyanakkor meglehetősen költséges. Sok esetben például az is segíti a keresési korlátok felszámolását, ha a közérdekű adatokat szolgáltató közigazgatási hierarchiát áttekintő honlap világos szerkezetű, jó keresési lehetőségeket biztosít, és minél több felhasználó számára ismert (vagyis a felhasználók tudják, hol keressék a keresési lehetőségeket). Kiegészítő adatszolgáltatási beruházásokkal valószínűleg

¹⁹⁰ Ezen feltehetően változtat az elektronikus információszabadság-törvény által előírt proaktív közzétételi kötelezettség és az elektronikus adatigénylés lehetősége.

jelentősen megnövelhetőek a nyilvánosságtól várt többlethasznok, ugyanakkor láttunk arra is példát a mobiltelefon-elérhetőségű gépjárműkereső szolgáltatás esetében, hogy nem minden kiegészítő intézkedés bevezetése hatékony.

Kutatói adatok esetén sokkal inkább valószínűsíthetjük, hogy az adatok nyilvánosságra hozatalával jelentős mértékben megnövekszik azok felhasználása, és pár év múlva a kutatói munka társadalmi hasznai is. A kutatói adatokat elemző fejezetben mutattunk is magyarországi példákat, hogy egy-egy adatforrás megnyitása hogyan növelte meg az adott területről felhalmozódó tudás mennyiségét. Ennek oka az, hogy a kutatók eleve rendelkeznek a hozzáférés technológiájával, valamint igen kicsik a nem pénzben mérhető keresési költségek is a szakmai információáramlás miatt. A kutatói adatok keresletét ugyanakkor jelentősen befolyásolja az adatok minősége, hogy azok mennyire tisztítottak, jól dokumentáltak, stb. Ezen a területen is indokoltak tehát kiegészítő adatkezelési beruházások, amelyek az egyébként minden adatkérő által egyénileg elvégzett feladatokat szükségtelenné teszik, így minden potenciális adatfelhasználó számára megtakarítást jelentenek.

A hazai adatkezelő intézmények gyakorlatát vizsgálva nem tudunk számszerű becsléseket adni arra nézve, hogy egy adott árváltozás hatása mekkora keresletváltozást jelent az adatszolgáltatás piacán. Nem véletlen, hogy ilyen elemzés nem készült, ugyanis maguk az intézmények sem igazán rendelkeznek a forgalomra vonatkozó adatokkal, vagy ha rendelkeznek is, nem bocsátották azokat rendelkezésünkre. Véleményünk szerint ez részben annak következménye, hogy a közérdekű adatszolgáltató szervezetekkel szemben a hatékony gazdálkodás nem valódi elvárás. Ennek következtében az intézmények nem érdekeltek abban, hogy teljesítményüket piaci módszerekkel mérijék. Amennyiben forgalmuk adatait ezek a szervezetek átláthatóan dokumentálnák, becsülhetővé válna az adott adatforrás iránti kereslet is. Ennek alapján konkrét előrejelzéseket is adhatnánk, hogy egyes felhasználói csoportok kereslete hogyan változna például az adatok árának drasztikus csökkentése esetén, illetve hogy ebből mekkora jóléti nyereségek származnak.

10.1.2.A kínálati oldal

Miként ezt a közgazdasági áttekintő fejezetben részletesen tárgyaltuk, az adatkezelő és -szolgáltató szervezetek speciális költségstruktúrával rendelkeznek: az adatszolgáltatás megfelelő szintjének biztosítása nagy fix költséggel jár, ugyanakkor egy már meglévő adat elhanyagolhatóan kis határköltséggel reprodukálható minden adatkérés alkalmával. Mivel a közérdekű adatok tömege ugyanakkor adott időpillanatban rendelkezésre áll, hajlamosak vagyunk azt mondani, hogy a fix költséget már megfizette a társadalom, vagyis az adatszolgáltatás árának meghatározásakor ezek megtérülését nem kell figyelembe vennünk, csak a kis határköltségeket.

Az adatszolgáltatás technológiája ugyanakkor sok esetben csak arra az adatigénylési kapacitásra készültek fel, amely az éppen aktuális árak mellett indokoltnak tűnik. Több adatszolgáltató (például az Országos Meteorológia Szolgálat) elbeszéléséből kitűnt, hogy a jelenlegi apparátus (szerverkapacitás, stb.) nem lenne képes kezelni a teljes adatbázisuk nyilvánosságra hozásának következtében előálló többletforgalmat. Azok az intézmények, amelyek közérdekű adataik jelentős részét már nyilvánossá tették (KSH), megerősítik ezt a nagy költségigényt. Amennyiben ezek a beruházások nem valósulnak meg, a lecsökkentett ár hiába teremtené nagyobb keresletet – azt a kínálat nem tudná kielégíteni.

A kielégítetlen túlkeresletnek két káros következménye is lehet. Egyrészt ha véletlen módon dől el, kinek az adatkérése teljesül, akkor semmi sem biztosítja, hogy azok jutnak az adathoz, akik azt a legtöbbre értékelik, míg a kereslet és kínálat egyensúlyában ez az úgynevezett elosztási hatékonyság biztosított lenne. Másrészt a szűkös kínálat egyes, a szolgáltatást többre értékelő szereplőket arra ösztönözhet, hogy jól irányzott juttatások révén férjenek hozzá az őket érdeklő adatokhoz, amelyből az intézmény egyes szereplői vagy információközvetítők

szereznek járadékokat.¹⁹¹ Bár ezen járadékvadászati elemek bizonyos értelemben megvalósítják az elosztási hatékonyságot, alkalmazásuk során olyan erőforrások vesznek kárba, amelyek más területen hatékonyan hasznosulhatnának. Összefoglalásképp, amennyiben célunk a közérdekű adatok nyilvánossága, meg kell teremteni a kiszolgáló technológia költségvetési forrásait, különben az adatnyilvánosságtól várt hasznok rövid távon sem realizálódnak teljes mértékben.

A ma rendelkezésre álló közérdekű adatok összessége ugyanakkor az idő előrehaladtával veszít hasznosságából, a jövőben aktuális döntések következményeinek felmérésénél ugyanis az éppen érvényben lévő adatok felhasználása az optimális. Ez hosszabb távon újabb adatok gyűjtésével és rendszerezésével jár, amelynek beruházási költségeit a szolgáltatónak ismét biztosítani kell. Bár továbbra is fennáll, hogy a közérdekű adatok többségét eredendően az államapparátus igényeinek kiszolgálására gyűjtik, vagyis ennek a költségnek a jó részét mindenképpen ki kell fizetni, azon adatkezelési (például adatvédelmi vagy a már korábban említett minőségjavító) eljárások forrását, amelyek révén ez az adattömeg a nyilvánosság elé kerül, továbbra is biztosítani kell.

Amennyiben ezek a költségvetési források rendelkezésre állnak, a keletkező hasznok és költségek összevetése után nem kérdés, hogy érdemes mindent megtenni a közérdekű adatokhoz való ingyenes hozzáférés biztosítására. Ha viszont a forrás léte nem nyilvánvaló, több út is kínálkozik. Az első lehetséges forgatókönyv szerint elvárhatjuk az intézményektől az ingyenes hozzáférés biztosítását, ugyanakkor nem biztosítunk számukra ahhoz elegendő forrást. Ennek a lehetőségnek a következményeit a fentiekben már ismertettük, és láttuk, hogy rövid távon ez hatékonyságveszteségekhez, hosszú távon pedig a szolgáltatás romlásához is vezethet.

A második lehetőség, hogy megengedjük a határköltségnél magasabb árak alkalmazását – ez (ha nyers adatokról van szó) már eltér az Avtv. előírásától.

¹⁹¹ Földhivatali adatkérésnél hosszú időn keresztül ismert a klasszikus várakozás, de kisebb költségű állampolgári út, illetve az „ügyvédem kapcsolataim” keresztül vezető gyors, de költségesebb út.

Amennyiben az árakat önköltség-alapon (vagy másképpen: átlagköltség-alapon) választják, mint ezt számos adatszolgáltató intézmény példáján láttuk is, akkor ez megteremti a fix költségek egészének vagy egy részének fedezetét, és csökkenti a költségvetés által biztosítandó forrásokat. Az árazás megengedésével ugyanakkor előtérbe kerül az adatszolgáltató ellenőrzésének szükségessége, a megteremtett bevételek ugyanis kisebb ösztönzőket adhatnak az intézmény költséghatékony működtetésére. A szigorúbb monitoring, hogy a választott árakban csak az indokolt költségeket érvényesítsék, többletköltségeket is jelent, amelyekkel számolnunk kell az árazási elvek esetleges bevezetésekor.

Az árazási elvek alkalmazása esetén további kérdés az egységes vagy diszkriminatív árak alkalmazása.¹⁹² Mint a közgazdasági fejezetben ezt megmutattuk, a különböző fogyasztói csoportok irányába alkalmazott egységes ár hatékonyságvesztéseket okozhat, ezért érdemes relatíve alacsonyabb árat alkalmazni azon adatigénylők felé, akik az árváltozás hatásaira érzékenyebben reagálnak, illetve adatfogyasztásuk nagyobb pozitív externális hatásokkal jár a társadalom többi tagja irányába. Erre az árazási alapelvre, miszerint állampolgárok vagy kutatók alacsonyabb áron jutnak hozzá ugyanahhoz az adathoz, mint a piaci szereplők, több adatszolgáltató gyakorlatában láttunk példát mind a klasszikus, mind a kutatói adatok esetében.

Az árdiszkrimináció ugyanakkor felveti azt a kérdést, hogy a csak magasabb árra jogosult szereplők nem juthatnának hozzá valamilyen módon az alacsonyabb áron a kívánt adatokhoz. Erre számos többé-kevésbé illegális módszer is kínálkozik, amely lehetőségek megszüntetése a szabályozó elemi érdeke, hiszen különben visszaáll a nem hatékony egységes ár esete. A kutatói adatokkal kapcsolatos nemzetközi példákban láttuk, hogy amennyiben kellően szigorú törvények vonatkoznak az adatokkal való visszaélés büntethetőségére, akkor az megnöveli a magasabb ár kikerülésének kockázatát. A törvényi környezet hatékonysága természetesen országonként erősen különbözik, ugyanakkor több esetben látunk arra példát, hogy az árazási (hozzáférési) elvek alkalmazhatósága javulhat attól, ha azokat más kiegészítő elemekkel kölcsönhatásban vizsgáljuk.

¹⁹² Itt ismét meg kell említenünk, hogy az Avtv. szellemében nem alkalmazhatóak eltérő árak különböző tulajdonságú vagy szándékú csoportok tagjaival szemben.

10.2. Az adatnyilvánosság intézményi adottságai

Mint láthattuk, a magyarországi adatkezelő intézmények egy része nem bizonyul egyértelműen együttműködőnek a birtokukban lévő adatokra vonatkozó adatkérések teljesítésében. Különösen azok az intézmények tanúsítanak nagyobb ellenállást az adatszolgáltatás irányában, ahol ez a feladat nem elsődleges tevékenységi körként definiált (lásd az APEH példáját). Erre való hivatkozva elvethetik az esetleges adatkérelmeket, különösen, ha az adatok kiadása számukra további költséget is jelent. Ezekben az esetekben igen hasznos lenne a jogértelmezés: megvilágítaná, hogy az adatszolgáltatási kötelezettség kiterjed minden, az intézmény kezelésében lévő adatra, hacsak az más elvbe nem ütközik.

Az adatvédelmi elvek különösen jól használhatók az adatkérés elhárítására. Az intézmények egyrészt azért kényszerülnek erre hivatkozni, mert amennyiben bármifajta jogsértés történik, akkor rájuk könnyen visszavezethető a felelősség. Ezt a kockázatkerülő viselkedést felerősíti, hogy az intézmények nincsenek teljesen tisztában azzal, milyen adatvédelmi elveket érvényesítve adhatnák ki (vagy egyáltalán kiadhatnák-e) az érzékenyebb adatokat. Ezen a téren is szükségszerűnek érezzük az alkalmazandó alapelvek világos, felvilágosító jellegű kifejtését – mit jelent például az a követendő elv, hogy az adott személy ne legyen egyértelműen beazonosítható? Véleményünk szerint mindenképpen tudatosítani kellene mind az intézményekben, mind az adatszolgáltatásról való közgondolkodásban, hogy az adatvédelem sok esetben pusztán többletköltségeket jelent, és nem leküzdhetetlen akadályt.

A legtöbb esetben az adatvédelemre támaszkodó egyértelmű kérelemmegtagadás mögött mégsem a lehetőségek nem ismerete, hanem az áll, hogy az adott adatbázis nem áll az adatkezelő rendelkezésére az adatvédelmi szabályoknak megfelelő formában. Ha tiszta formájában érvényesítjük az Avtv. árazásra vonatkozó ajánlását, akkor az adatszolgáltató nem számíthatja fel az adatbázisok

anonimizálásának vagy összekapcsolásának fix költségeit, ami ismét ellenérdekeltséget teremt ezen típusú kérelmek teljesítésére.

A kutatási adatokat kezelő és szolgáltató intézményeknél gyakran az az aggály motiválja az adatszolgáltatás költségmentessé tételével szembeni ellenállást, hogy a nyilvános adatoknak ezután bárki szabad értelmezést adhat, ezzel egyrészt kisajátítva magának az eredményeket, másrészt az esetleges félreértelmezések továbbadásával károkat is okozhat. Az első problémára megoldást adhat az a (gyakorlatban egyre inkább elterjedő) kiegészítő intézkedés, amely kötelezi az adatok felhasználóit, hogy az adatok forrásaként mindig megjelöljék a felhasznált adatok eredetét, növelve ezzel az adatszolgáltató reputációját is. Amennyiben az adatértelmezőket kötelezik arra, hogy közöljék a következtetéseikhez használt pontos módszertant, az megoldást jelent a második problémára is, mivel a nyilvános adatok és az elemzési módszerek ismeretében az elvégzett kutatás reprodukálható.

A harmadik elem, ami az adatszolgáltatókat elriasztani látszik adataik nyilvánosságra hozatalától, hogy adatkezelési módszereik nyilvánosságra kerülnek és a forráshiány okozta adatkezelési hiányosságai „napvilágra kerülése” esetleg negatív színben tünteti fel őket. Véleményünk szerint itt szemléletbeli változásra és az adatok minőségének egyszeri, a nyilvánosság biztosítását megelőző javítására van szükség, a megfelelő anyagi támogatással. Ha egy adat jelenleg nem megfelelő minőségű, annak sok esetben az az oka, hogy pillanatnyi, az intézmény által szem előtt tartott hasznosítási módja nem követel meg jobb minőséget. A más, nagyobb precizitást igénylő felhasználási mód létét az intézménynek egyszer kell beépítenie munkamódszerébe. Tudatosítani szükséges azt is, hogy az adatok nem tökéletes minősége az adatok előállításának technológiai velejárója, azt eleve kiküszöbölni nem lehet, csak javítani.

Az adatokat felhasználó más intézmények vagy kutatók tevékenysége ugyanakkor számos pozitív visszahatást eredményezhet, amely jelentkezhet akár maguknak az adatoknak a minőségi javításában, illetve az intézmény működésére vonatkozó

hasznos javaslatokban is. Ha tehát rövid távon a nyilvánosságra hozás kisebb hátrányokkal jár, hosszú távon ezek az áldozatok már megtérülhetnek. Ezt ugyanakkor az adatszolgáltató intézménnyel hitelesen érzékeltetni kell, hiszen az ő együttműködése kulcsfontosságú ebben a folyamatban.

10.3. Ajánlások

Ajánlásaink alapvetően két területre koncentrálnak. Az egyik az árazási és az ehhez kapcsolódó finanszírozási kérdésekre vonatkozik, a másik az adatok hozzáféréseinek módosítására. Az ajánlások között vannak, amelyek önmagukban is megvalósíthatók, ezeket külön hangsúlyozzuk. Túlnyomó többségük azonban csak a többivel együtt, azokkal összefüggésben értelmezhető.

Általánosan elterjedt kapcsolódási felület a klasszikus közérdekű adatokhoz. Elsőként visszatérnénk ahhoz a klasszikus adatokkal kapcsolatos eredményhez, hogy a hozzáférés költsége nem egyenlő a pénzbeli költségekkel. Különösen fontosak azok a (jórészt tudásbeli) fix költségek, amelyek bizonyos adatelérési formák, például az internet használatához szükségesek: a magyar lakosságnak csak kisebbsége képes ezt a médiumot rutinszerűen használni. Paradox módon éppen az Internet az a csatorna, amelyen az erre kötelezett intézmények a legtöbb közérdekű adatot elérhetővé teszik. Meg kell tehát fontolni, legalább a hazai Internet-használat szintjének emelkedéséig, olyan központosított elérési felület létrehozását, amely az Internetnél szélesebb körben elterjedt. Véleményünk szerint a telefon a legalkalmasabb médium erre a célra. Az elképzelt helyzetben az egyes adatgazdák valóban csak az interneten teszik elérhetővé az adataikat, de működik egy olyan *call-center*, amelynek munkatársai a hívásokat fogadva az így elérhető adatokban keresni tudnak, helyi hívás díjáért felvilágosítást adva az érdeklődőknek. A rendszer működése a mindenki számára ismert különleges tudakozóhoz hasonlítana. Kutatásunk rámutatott, hogy az adatok léte sok esetben nem nyilvánvaló az érintettek számára. Egy ilyen rendszer létezését több állami intézményben, önkormányzatban, orvosi rendelőben, okmányirodában hirdetni kellene. Bár az ilyen elérési lehetőség kialakításakor csábító eleve nagy kapacitást

tervezni, például a gépjármű-nyilvántartás esete intő példával szolgál: inkább a kereslet felmérése után érdemes bővíteni azt, mert a kereslet meglepően kicsi lehet.

Differenciált, de szabályozott árazás lehetősége. Az eddigi elemzésünkéből kitűnt, hogy bár a költségmentes hozzáférhetőség megvalósítását hosszú távon elérendő célnak gondoljuk, kialakítására nem minden esetben vannak meg az anyagi feltételek. Ezalatt az átmeneti időszak alatt inkább érdemes megengedni az érintett intézményeknek, hogy bizonyos szolgáltatásaikért a határköltségnél magasabb díjat számoljanak fel, minthogy ingyenességre kényszerítve őket veszélyeztessük a szolgáltatás színvonalát.

Erre az intézkedésre ott van mindenekelőtt szükség, ahol az adatkezelő szakértői vagy kutatási célú adatokat szolgáltat. A klasszikus közérdekű adatok esetében, ahol az adat már alapvető formájában hasznosítható, már ma is szinte minden esetben kialakult differenciálódás. Ennek eredményeként az adatok legegyszerűbb, még kielégítő formája – az Avtv. rendelkezésével egyezően – ingyenesen elérhető, értéknövelt változata azonban térítésköteles piaci termék.

Mint az egyes magyarországi példákon láttuk, a mai magyar gyakorlat elnézőnek tűnik ebből a szempontból, az intézményeket árképzésébe nem igazán avatkozik bele, miközben elvileg megköveteli a teljes nyilvánosságot. A szabályozatlan teljes nyilvánosság ezen formájánál előnyösebbnek gondoljuk a szabályozott, de célszerűen korlátozott nyilvánosságot. Ebben a jogszabály elismerné, hogy bizonyos költségek elszámolhatóak, de árszabályozott körülmények között, így érvényesíthetők lennének olyan szempontok is, amelyeket az intézményeknek nem áll érdekükben követni (például és főként a keletkező externáliák). A szabályozás fontosságát az is indokolja, hogy hiányában az adatgazdák a sok esetben természetesen, vagy pusztán az állami finanszírozás kizárólagos címzettjeként megszerzett monopolhelyzetükkel visszaélhetnek.

Kétrészes árazás alkalmazásának megfontolása. Mint láttuk, csak egységárral nem kezelhető egyszerre optimálisan a bevételszerzés és az elosztási hatékonyság érvényesítése. A közhasznú adatokat nagy tömegben igénylő professzionális felhasználók (például média vagy nagyobb kutatóintézetek) irányában érdemes lenne ezért megengedni licenz-szerződések alkalmazását, amelyek egy fix belépési díj ellenében valóban ingyenes és korlátlan elérhetőséget biztosítanak a számukra szükséges adatokhoz.

Az alapkompétencia keretében termelt adatok körének explicit kijelölése és az értéknövelt szolgáltatások nyilvánossága. A differenciált árazás megengedésének javasolt módja feltételezi, hogy meg tudjuk mondani, egy adott intézmény esetében mi az alaptevékenységeként keletkező adat, és mi az, ami ezen felül keletkezik. Ennek fontosságát mutatja, hogy számos intézmény állít elő ma is „értéknövelt” szolgáltatásként származtatott adatokat. Az alapkompétencia meghatározása azért fontos, ne történhessen meg, hogy adatgazdák értéknövelt szolgáltatássá minősítik alapvető adataikat. Az adatok nyilvánossága terjedjen ki az értéknövelt szolgáltatások körére is, hiszen ex post azok is közérdekű adatok. Az üzleti titok védelme érdekében ezek az adatok azonban csak a szolgáltatás után bizonyos idővel legyenek költségmentesen elérhetőek mindenki számára.

A tudományos kutatás társadalmi hasznának kodifikálása és tudatosítása. Szükséges lenne a jogi szabályozásban megjeleníteni azt a gondolatot, hogy az állami működés hatékonyságának ellenőrzéséhez értékelő kutatásokat szükséges végezni, amelyek elősegítéséhez minden, az állam birtokában levő adat rendelkezésre kell, hogy álljon, amennyiben az személyes jellegétől megfosztható. Több olyan adatforrás létezik, amelyet az adatkezelő eredendően személyes adatként kezel, de személyességétől egyszerű művelettel megfosztható. A jelenlegi szabályozás a személyes adatok kezelését jogszabályi megokoláshoz köti, ami könnyű kibúvót jelent az adminisztratív adatok kutatási célra való kiadása és összekapcsolása alól, hiszen azok adott formájukban személyes adatok, amelyeket első lépésben kezelni, anonimizálni kell. Ezt a problémát a fenti cél általános rögzítése megoldaná. A jogi szabályozás mellett kívánatos az, hogy az adatokat

kezelő intézmények legalább vezető pozícióban levő munkatársait felvilágosítsák az elemző munka jellegéről, fontosságáról és adatigényéről.

Amelyik intézmény adatot gyűjt, legyen kötelezően adatszolgáltató.

Egyértelművé kellene tenni az Avtv.-ben és szükség szerint más jogszabályokban azt a gondolatot, hogy minden olyan állami szervezetnek és hivatalnak elsődleges feladata az adatok szolgáltatása, amely adatokat rendezett formában és számítógéppel kezel. Ezzel a jelenlegi szabályozás lényegében nem módosul, csak hangsúlyt kap az adatszolgáltatás fontossága.

Ne csak a közérdekű, hanem a közérdekűvé tehető adatok szolgáltatása is legyen kötelező.

Az adatok anonimizálását az adatkezelők számára kötelezővé lehetne tenni minden olyan esetben, amikor az technikailag megtehető. Az adatszolgáltatási kötelezettség terjedjen ki ezekre az adatokra is úgy, mintha már rendelkezésre állna anonimizált változatuk. Természetesen ez a művelet, hasonlóan az adatszolgáltatáshoz, térítésköteles lehetne, hiszen az anonimizálás költsége itt végeredményben az adatok hozzáféréseinek költsége. Mivel az anonimizálást egy adatbázison csak egyszer kell elvégezni, annak díját a várható kereslet megbecsülése után átlagköltség-alapon érdemes megállapítani.

Metaadatok. Az adatok értelmezésének feltétele az, hogy a felhasználó tisztában legyen az adat keletkezésének folyamatával. A közérdekű adatok nyilvánossága mellett legyen kötelező a rájuk vonatkozó metaadatok részletes és korlátozás nélküli nyilvánosságra hozatala még akkor is, ha egy adott adatforráshoz a hozzáférés valamilyen oknál fogva a felhasználók valamely csoportjára vonatkozólag korlátozott. Legyen támogatott az adatkezelő intézményeknél valamely minőségbiztosítási rendszer alkalmazása, és a rendszer dokumentációja legyen elérhető.¹⁹³

Adatbázisok megfelelő körülmények közötti összekapcsolásának támogatása. Az eddig javasolt módosítások után az adminisztratív adatok továbbra

¹⁹³ Az adatok egy jól körülhatárolt része tekintetében az elektronikus információszabadság-szabályozás ezt megoldhatja.

sem lennének összekapcsolhatók, még akkor sem, ha tökéletesen anonimizálhatók. Meg kellene teremteni annak a jogi háttérét, hogy a Belügyminisztérium Központi Nyilvántartó és Választási Hivatala köteles legyen a TAJ és adószámokat összekapcsolt kiadását nem igénylő, de a TAJ és adószámmal azonosított adatbázisok összekapcsolásához a Hivatal szolgáltatását igénylő kérés teljesítésére. Amint ezt a hatodik fejezet példájában is bemutattuk, ez a művelet technikailag az adatvédelmi elvek legteljesebb figyelembevételével elvégezhető, csak a BMKÖANYVH-ra vonatkozó jelenlegi szabályozás nem teszi lehetővé.

Kutatószobák létesítése és jogi háttér biztosítása az érzékeny adatok elérésére. Vannak olyan esetek, amikor az adatok anonimizálása nem megoldható, például azért, mert a feltett kérdés lényege olyan, ami személyes adatok kezelését igényli (bár nem egyedi adatokból von le következtetéseket). Ilyenkor arra van szükség, hogy a kérdést vizsgáló kutató az adatot kezelő hivatal munkatársaival megegyező jogosultságot kapjon az adatok kezelésére – ez a nemzetközi tapasztalatok alapján bejáratott gyakorlatnak tekinthető. Ennél a műveletnél azonban nemcsak az anonimizálás nehézségei és bizonytalanságai merülnek fel, ami kiváltja az adatgazdák ellenállását, de nincs meg a kutatót az intézmény munkatársainak hozzáférési szintjére emelő jogszabályi háttér sem. Szükség van tehát ennek kidolgozására és a joganyagba iktatására. Emellett az anonimizálási kényszerrel párhuzamos kötelezettséget kell teremteni arra, hogy az ilyen jellegű adattal rendelkező intézmények kutatószobát hozzanak létre az érzékeny adatok kezelésére. Mivel a kutatószoba létesítése jelentős beruházás, így a kötelezettség nem lehet általános – az érintett intézmények körét szakértőknek kell kiválasztani, minimális programként érintett a KSH.

Jogértelmezés. Tapasztalatunk szerint a jogi következményektől való félelem sok adatkezelőt visszariasztja adatainak közreadásától. A tapasztalat azt mutatja, hogy a jogszabály általános érvénye miatt a konkrét esetekben sokszor nem nyilvánvaló, mit akar mondani. Ez gyakran az adatkezelő érdekeit legjobban szolgáló értelmezéshez, így az adatok kiadásának megtagadásához vezet. Segíthetne ezen a helyzeten, ha a kötelezettség megteremtése mellett az adatkezelők hiteles forrásból

gyakorlati útmutatást is kapnának kidolgozott adatkezelési minták formájában. Az adatkérések terén kisszámú kiemelt eset van: ezeknél konkrét előzetes jogértelmezésre lenne szükség.

Az adatvédelmi biztos szerepe. Az adatvédelmi biztos kulcsszereplő az adatok hozzáférhetőségében. Jelenleg tapasztalható működésének alapja érezhetően annak felkutatása, hogy mi lehet egy adat visszatartásának indoka, mi akadályozhatja annak nyilvánosságát. A kívánatos az lenne, ha a biztos szakértelmét az adatfelhasználás szolgálatába állítaná, az általa feltett fő kérdés az lenne: hogyan lehet megoldani, hogy az adott adat felhasználható legyen. Megfelelő jogszabályi garancia és kötelezettség, a biztos feladatainak a javasolt szemponttal való bővítése egyértelműbbé tenné a helyzetet és megszüntetné a személyes megfontolás jelenleg érezhető hatását.

Diszkrimináció lehetősége a nem szakértő felhasználók ellen. A félelem másik fő forrása az, hogy alapvető adatokhoz, melyekből az érintett intézmény származtatott adatokat készít, nem kellőképpen felkészült felhasználók is hozzáférést nyernek. Ez a félelem mind elméleti alapon, mind a nemzetközi gyakorlat alapján (ahol erőteljesen védekeznek az ilyen jellegű felhasználás ellen) jogosnak mondható, az Avtv. azonban jelenlegi formában nem teszi lehetővé azt, hogy közérdekű adatok hozzáféréséből bárki is kizárható legyen. Az intézmények nem kis részben azért nem teszik adatvédelmi szempontból hozzáférhetővé (nem anonimizálják) értékes mikroadataikat, mert így mentesülhetnek a kötelező nyilvánosság elől. Ha egy esetleges anonimizálási kötelezettség mellett megmarad az Avtv. diszkriminációt tiltó rendelkezése, az az adatkezelő intézményeket és ezáltal felhasználóikat is rendkívül kellemetlen helyzetbe hozza. Feltételezhető egyúttal, hogy az adatkezelő intézmények elemi érdekeik védelme érdekében sabotálnák az erőviszonyok ilyen egyoldalú megváltoztatását. Éppen ezért azt javasoljuk, hogy a nemzetközi gyakorlatnak megfelelően legyen megengedhető az elemi adatok adathozzáféréseinek korlátozása, megfelelő háttérhez kötése (például tudományos intézmény és elismert szakemberek garanciavállalása). A teljes nyilvánosság is megadható, de csak az adat pillanatnyi relevanciájának, így a

félreértelmezésből fakadó nagy veszteség veszélyének elmúltával, ami a vizsgált esetekben néhány évre tehető. Társadalmi szinten így megvalósul az adatok nyilvánossága, az intézmények jogos félelme pedig orvosolható.

11. Felhasznált és hivatkozott források jegyzéke

2. fejezet

Richard Schmalensee (1981): Output and Welfare Implications of Monopolistic Third-Degree Price Discrimination, *American Economic Review* 71, 242-247.

Hal Varian (1985): Price Discrimination and Social Welfare, *American Economic Review* 75, 870-75.

Hal Varian (2001): *Mikroökonómia középfolon*, Közgazdasági és Jogi Kiadó, Budapest.

3. fejezet

Eötvös Károly Intézet (2003): Elektronikus adatnyilvánosság.

George Stigler (1961): Az információszerzés közgazdaságtana, magyarul megjelent in: George Stigler: Piac és állami szabályozás (válogatott tanulmányok), 1989, Közgazdasági és Jogi Kiadó, Budapest, 232-253.

4. fejezet

John Abowd és Julia Lane (2003): Economics of Data Confidentiality, kézirat.

George Duncan (1999): Innovative Solutions in Providing Access to Micro Data, kézirat.

George Duncan és szerzőtársai (2001): Disclosure Limitation methods and information loss for tabular data, in: Confidentiality, Disclosure and Data Access: Theory and Practical Applications for Statistical Agencies, szerkesztette Peter Doyle és társai, North Holland, Amsterdam.

Kertesi Gábor (2000): A cigány foglalkoztatás leépülése és szerkezeti átalakulása 1984 és 1994 között. Közgazdasági Szemle 57, 406-443.

National Research Council (2005): Expanding Access to Research Data: Reconciling Risks and Opportunities, Panel on Data Access for Research Purposes, National Academies Press

Bódis L. – Galasi P. – Micklewright J. – Nagy Gy. (2005): Munkanélküli-ellátás és hatásvizsgálatai Magyarországon. MTA Közgazdaságtudományi Intézet, Budapest

Kapitány Zs. – Molnár Gy. – Virág I. (2005): Háztartások a tudás- és munkapiacra.
MTA Közgazdaságtudományi Intézet, Budapest

Halpern L. – Kertesi G. – Koren M. – Köllő J. – Kőrösi G. – Vincze J. (2004): A minimálbér költségvetési hatásai. MTA Közgazdaságtudományi Intézet
Műhelytanulmányok MT-DP 2004/4

Magyar Statisztikai évkönyv 2003. Központi Statisztikai Hivatal, Budapest

Fazekas K. – Varga J. (szerk.) (2004): Munkaerőpiaci Tükör 2004. MTA
Közgazdaságtudományi Intézet, Budapest

5. fejezet

Az északi országokkal kapcsolatban

Access to Microdata in the Nordic Countries – az északi országok statisztikai
hivatalainak közös kiadványa, Statistics Sweden, SCB-Tryck, Örebro, 2003

Access to Microdata From Statistics Denmark, kézirat.

Az Egyesült Államokkal kapcsolatban

Census Bureau: <http://www.census.gov/>

Adatvédelemre vonatkozó törvényidézetek a CB oldalain:

http://www.census.gov/privacy/files/data_protection/004032.html

Adatvédelemmel kapcsolatos tapasztalatok a CB-nál:

<http://www.census.gov/srd/sdc/microdataprotection.pdf>

BLS: <http://www.bls.gov/bls/infohome.htm>

Adathozzáférésekről a BLS-nél:

James Spletzer (2003): Microdata Access at the U.S. Bureau of Labor Statistics,
kézirat, elérhető: <http://www.micro2122.scb.se/papers.asp>

NBER és az adatok: http://www.nber.org/cgi-bin/get_bar.pl?bar=data

PSID: <http://psidonline.isr.umich.edu/>

HRS: <http://hrsonline.isr.umich.edu/>

SIPP: <http://www.sipp.census.gov/sipp/>

FOIA: <http://www.usdoj.gov/04foia/foiastat.htm>

Privacy Act: <http://www.usdoj.gov/04foia/privstat.htm>

CIPSEA: <http://scitech.dot.gov/research/human/docs/hfcc/title-v.pdf>

Adatok preparálásáról:

Julia Lane, John Abowd és Simon Woodcock (2005): [Proposed Session on Masked Data and Confidentiality Protection](#), kézirat, elérhető: <http://www.micro2122.scb.se/papers.asp>

Nagy-Britanniával kapcsolatban

ONS: <http://www.statistics.gov.uk/about/>

ESRC: <http://www.esrcsocietytoday.ac.uk/ESRCInfoCentre/about/>

ISER: <http://www.iser.essex.ac.uk/>

ESDS: <http://www.esds.ac.uk/>

UKDA: <http://www.data-archive.ac.uk/>

MIMAS: <http://www.mimas.ac.uk/about.html>

Adatok a MIMAS oldalain:

<http://www.mimas.ac.uk/services/primarydata.html>

Az Athens azonosítóról: <http://www.athens.ac.uk/>

Adatvédelmi szabályozások az Egyesült Királyságban:

<http://www.informationcommissioner.gov.uk/eventual.aspx?id=34>

Szabad információáramlás az Egyesült Királyságban:

<http://www.informationcommissioner.gov.uk/eventual.aspx?id=33>

National Statistics Code of Practice:

http://www.statistics.gov.uk/about/national_statistics/cop/default.asp

Az Európai Unió szabályozásáról

Recent EU legislation for research access to confidential data – implementation and implications, kézirat, Conference of European Statisticians 2003

5. fejezet

H.L. (2005): „Interjú dr. Pukli Péterrel”. *Statisztikai szemle*, 83. évfolyam 1. szám. 2005. január